Cerebral Cortex, 2017; 1–12

doi: 10.1093/cercor/bhx277
Original Article

ORIGINAL ARTICLE

Neural Encoding of Auditory Features during Music Perception and Imagery

Stephanie Martin[1,2], Christian Mikutta[2,3,4], Matthew K. Leonard[5], Dylan Hungate[5], Stefan Koelsch[6], Shihab Shamma[7,8], Edward F. Chang[5], José del R. Millán[1], Robert T. Knight[2,9] and Brian N. Pasley[2]

[1]Defitech Chair in Brain-Machine Interface, Center for Neuroprosthetics, Ecole Polytechnique Fédérale de Lausanne, Lausanne 1015, Switzerland, [2]Helen Wills Neuroscience Institute, University of California, Berkeley, CA 94720, USA, [3]Translational Research Center and Division of Clinical Research Support, Psychiatric Services University of Bern (UPD), University Hospital of Psychiatry, Bern 3072, Switzerland, [4]Department of Neurology, Inselspital, Bern, University Hospital, University of Bern, Bern 3010, Switzerland, [5]Department of Neurological Surgery, Department of Physiology, and Center for Integrative Neuroscience, University of California, San Francisco, CA 94143, USA., [6]Languages of Emotion, Freie Universitä, Berlin 14195, Germany, [7]Département d'études cognitives, École normale supérieure, PSL Research University, Paris 75006, France, [8]Electrical and Computer Engineering & Institute for Systems Research, Univ. of Maryland in College Park, MD 20742, USA and [9]Department of Psychology, University of California, Berkeley, CA 94720, USA

Address correspondence to Brian Pasley, Helen Wills Neuroscience Institute, University of California, 210 Barker Hall, Berkeley, CA 94720, USA. Email: bpasley@berkeley.edu

Abstract

Despite many behavioral and neuroimaging investigations, it remains unclear how the human cortex represents spectrotemporal sound features during auditory imagery, and how this representation compares to auditory perception. To assess this, we recorded electrocorticographic signals from an epileptic patient with proficient music ability in 2 conditions. First, the participant played 2 piano pieces on an electronic piano with the sound volume of the digital keyboard on. Second, the participant replayed the same piano pieces, but without auditory feedback, and the participant was asked to imagine hearing the music in his mind. In both conditions, the sound output of the keyboard was recorded, thus allowing precise time-locking between the neural activity and the spectrotemporal content of the music imagery. This novel task design provided a unique opportunity to apply receptive field modeling techniques to quantitatively study neural encoding during auditory mental imagery. In both conditions, we built encoding models to predict high gamma neural activity (70–150 Hz) from the spectrogram representation of the recorded sound. We found robust spectrotemporal receptive fields during auditory imagery with substantial, but not complete overlap in frequency tuning and cortical location compared to receptive fields measured during auditory perception.

Key words: auditory cortex, electrocorticography, frequency tuning, spectrotemporal receptive fields

Introduction

Auditory imagery is defined here as the mental representation of sound perception in the absence of external auditory stimulation. The experience of auditory imagery is common, such as when a song runs continually through someone's mind. On an advanced level, professional musicians are able to imagine the sound of a piece of music by looking at its printed notes (Meister et al. 2004). Behavioral studies have shown that structural and temporal properties of auditory features (see (Hubbard 2010) for complete review), such as pitch (Halpern 1989), timbre (Pitt and Crowder 1992; Halpern et al. 2004), loudness (Intons-Peterson 1992), and rhythm (Halpern 1988) are preserved during auditory imagery. Despite numerous behavioral and neuroimaging studies, it remains unclear how these auditory features are encoded in the brain during imagery. Experimental investigation is difficult due to the lack of observable stimulus or behavioral markers during auditory imagery. Using a novel experimental paradigm to synchronize auditory imagery events to neural activity, we quantitatively investigated the neural representation of spectrotemporal auditory features during auditory imagery in an epileptic patient with proficient music abilities.

Previous studies have identified anatomical regions active during auditory imagery (Kosslyn et al. 2001), and how they compare to actual auditory perception. For instance, lesion (Zatorre and Halpern 1993) and brain imaging studies (Zatorre et al. 1996; Griffiths 1999; Halpern and Zatorre 1999; Rauschecker 2001; Halpern et al. 2004; Kraemer et al. 2005) have confirmed the involvement of bilateral temporal lobe regions during auditory imagery (see Zatorre and Halpern 2005, for a review). Brain areas consistently activated with fMRI during auditory imagery include the secondary auditory cortex (Griffiths 1999; Kraemer et al. 2005; Zatorre et al. 2009), the frontal cortex, the sylvian parietal temporal area (Hickok et al. 2003), ventrolateral and dorsolateral cortices (Meyer et al. 2007), and the supplementary motor area (Mikumo 1994; Petsche et al. 1996; Halpern and Zatorre 1999; Halpern 2001; Schürmann et al. 2002; Brodsky et al. 2003). Anatomical regions active during auditory imagery have been compared to actual auditory perception to understand the interactions between externally and internally driven cortical processes. Several studies showed that auditory imagery has substantial, but not complete overlap in brain areas with music perception (Kosslyn et al. 2001)—for example, the secondary auditory cortex is consistently activated during music imagery and perception while the primary auditory areas appear to be activated solely during auditory perception (Griffiths 1999; Yoo et al. 2001; Halpern et al. 2004; Bunzeck et al. 2005).

These studies have helped to unravel anatomical brain areas involved in auditory perception and imagery; however, there is lack of evidence for the representation of specific acoustic features in the human cortex during auditory imagery. It remains a challenge to investigate neural processing during internal subjective experience like music imagery, due to the difficulty in time-locking brain activity to a measurable stimulus during auditory imagery. To address this issue, we recorded electrocorticographic neural signals (ECoG) of a proficient piano player in a novel task design that permitted robust marking of the spectrotemporal content of the intended music imagery to neural activity—thus allowing us to investigate specific auditory features during auditory imagery. In the first condition, the participant played an electronic piano with the sound output turned on. In this condition, the sound was played out loud through speakers at a comfortable sound volume that allowed auditory feedback (perception condition). In the second condition, the participant played the electronic piano with the speakers turned off, and instead imagined the corresponding music in his mind (imagery condition). In both conditions, the sound output of the keyboard was recorded. This provided a measurable record of the content and timing of the participant's music imagery when the speakers of the keyboard were turned off and he did not hear the music. This task design allowed precise temporal alignment between the recorded neural activity and spectrogram representations of music perception and imagery—providing a unique opportunity to apply receptive field modeling techniques to quantitatively study neural encoding during auditory imagery.

A well-established role of the early auditory system is to decompose complex sounds into their component frequencies (Aertsen et al. 1981; Eggermont et al. 1983; Tian 2004), giving rise to tonotopic maps in the auditory cortex (see Saenz and Langers 2014, for a review). Auditory perception has been extensively studied in animal models and humans using spectrotemporal receptive field (STRFs) analysis (Aertsen et al. 1981; Clopton and Backoff 1991; Theunissen et al. 2000; Chi et al. 2005; Pasley et al. 2012), which identifies the time-frequency stimulus features encoded by a neuron or population of neurons. STRFs are consistently observed during auditory perception tasks, but the existence of STRFs during auditory imagery is unclear due to the experimental challenges associated with synchronizing neural activity and the imagined stimulus. To characterize and compare the spectrotemporal tuning properties during auditory imagery and perception, we fitted 2 encoding models on data collected from the perception and imagery conditions. In this case, encoding models describe the linear mapping between a given auditory stimulus representation and its corresponding brain response. For instance, encoding models have revealed the neural tuning properties of various speech features, such as acoustic, phonetic, and semantic representations (Pasley et al. 2012; Tankus et al. 2012; Mesgarani et al. 2014; Lotte et al. 2015; Huth et al. 2016).

In this study, the neural representation of music perception and imagery was quantified by STRFs that predict high gamma (HG; 70–150 Hz) neural activity. High gamma correlates with the spiking activity of the underlying neuronal ensemble (Miller et al. 2007; Boonstra et al. 2009; Lachaux et al. 2012) and reliably tracks speech and music features in auditory and motor cortex (Crone et al. 2001; Towle et al. 2008; Llorens et al. 2011; Pasley et al. 2012; Sturm et al. 2014). Results demonstrated the presence of robustly measureable spectrotemporal receptive fields during auditory imagery with extensive overlap in frequency tuning and cortical location compared to receptive fields measured during auditory perception. Predictive accuracy was compared to alternative encoding and decoding models designed to control for potential motor confounds associated with piano playing. These results provide a quantitative characterization of the shared neural representation underlying auditory perception and the subjective experience of auditory imagery.

Materials and Methods

Participant and Data Acquisition

Electrocorticographic (ECoG) recording was obtained using subdural electrode arrays implanted in one patient undergoing neurosurgical treatment for refractory epilepsy. The participant gave his written informed consent prior to surgery and experimental testing. The experimental protocol was approved by the

University of California, San Francisco and Berkeley Institutional Review Boards and Committees on Human Research. The participant was a proficient piano player (age of start: 7, years of music education: 10, hours of training per week: 5). Prior studies have shown that music training is associated with improved auditory imagery ability, such as pitch and temporal acuity (Aleman et al. 2000; Lotze et al. 2003; Janata and Paroo 2006). A 256-electrode grid was implanted on the left hemisphere (Fig. 2A). Grid location was defined solely by clinical requirements. Inter-electrode spacing (center-to-center) was 4 mm and the electrode contact area diameter was 2.3 mm. Localization and co-registration of electrodes was performed using the structural MRI. Multi-electrode ECoG data were amplified and digitally recorded with sampling rate of 3052 Hz. ECoG signals were re-referenced to a common average after removal of electrodes with epileptic artifacts or excessive noise (including broadband electromagnetic noise from hospital equipment or poor contact with the cortical surface). In addition to the ECoG signals, the audio output of the piano was recorded along with the multi-electrode ECoG data.

Experimental Paradigm

The recording session included 2 conditions. In the first task, the participant played 2 music pieces on an electronic piano with the speakers of the digital keyboard turned on (perception condition; Fig. 1A). In the second task, the participant played the 2 same piano pieces, but the volume of the speaker system was turned off to establish a silent room. The participant was asked to imagine hearing the corresponding music in his mind as he played the piano (imagery condition; Fig. 1B). In both conditions, the audio signal from the line output jack of the keyboard was analogously recorded in synchrony with the ECoG signal at 24 414 Hz. The recorded sound allowed synchronizing the auditory spectrotemporal patterns of the imagined music and the neural activity when the speaker system was turned off and no audible sounds were recorded in the room. Figure 1B illustrates that the sound recorded in synchrony with the ECoG data (even when the speakers were turned off and the participant did not hear the music).

The 2 music pieces were the "Prelude in C Minor, Op. 28, No. 20" by Frederic Francois Chopin and the "Fugue No.1 In C Major, BWV 846" by Johann Sebastian Bach. Prior to the surgery, the imagery experiment was described to the patient who then selected these 2 music pieces based on familiarity and ease of performing the perception and imagery tasks. The patient reported that he played by reading the scores and was able to use auditory imagery vividly in synchrony with the spectral and temporal features elicited by the keypresses, including the lowest and highest pitches in the music pieces. In addition, self-report by the patient indicated he did not purposefully vary the tempo while playing, and a control encoding model analysis based on tempo variance revealed no significant relationship between tempo variance and the neural response (see Supplementary Materials for details). To assess general auditory imagery abilities, the participant completed the "Bucknell Audiotry Imagery Scale—Vividness and Control" (BAIS; Halpern 2015), a self-report assessment including subscales for musical, verbal, and environmental sounds that is based on 2 14-item questionnaires, respectively. In the Vividness test, the participant was asked to construct an auditory image (e.g., the sound of gentle rain), and rate his image for each item on a 7-point scale (1 = no image present at all; 7 = as vivid as actual sound). In the Control test, the original item was again described (e.g., the sound of gentle rain), and the participant was asked to change it to a new sound (e.g., the gentle rain turns into a violent thunderstorm), and rate how easily he could change the first image to the second image (1 = no image present at all; 7 = extremely easy to change the item). The participant's rating was 4.9/7 for the Vividness scale and 5.2/7 for the Control scale, which is within the range of previously reported assessments (2.9–6.9; Halpern 2015; Lima et al. 2015), suggesting that the participant was able to successfully perform the imagery task.

Feature Extraction

We extracted the ECoG signal in the high gamma frequency band from 8 bandpass filters (hamming window non-causal filter of order 20, logarithmically increasing center frequencies (70–150 Hz) and semi-logarithmically increasing bandwidths), and extracted the envelope using the Hilbert transform. Prior to model fitting, the power was averaged across these 8 bands, downsampled to 100 Hz and z-scored.

Auditory Spectrogram Representation

The auditory spectrogram representation was a time-varying representation of the amplitude envelope at acoustic frequencies logarithmically spaced between 200 and 7000 Hz. This representation was calculated by affine wavelet transforms of the sound waveform (output of the keyboard) using auditory filter banks that mimics neural processing in the human auditory periphery (Chi et al. 2005). To compute these acoustic representations, we used the NSL MATLAB toolbox (http://www. isr.umd.edu/Labs/NSL/Software.htm).

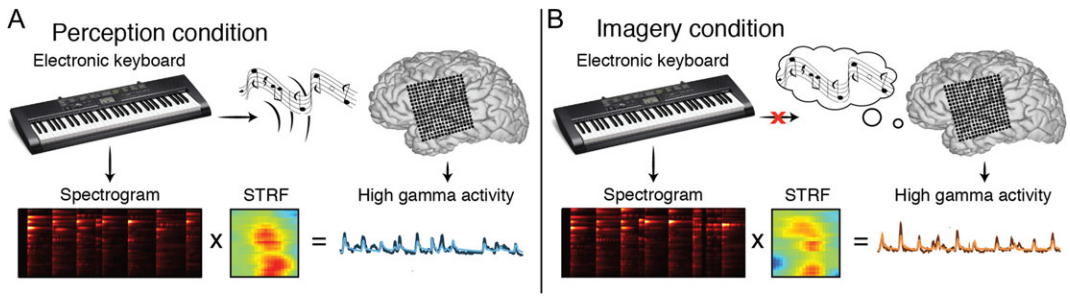

Figure 1. Experimental task design. (A) The participant played an electronic piano with the sound of the digital keyboard turned on (perception condition). (B) In the second condition, the participant played the piano with the sound turned off and instead imagined the corresponding music in his mind (imagery condition). In both conditions, the sound output of keyboard was recorded in synchrony with the neural signals (even when the participant did not hear any sound in the imagery condition). The models take as input a spectrogram consisting of time-varying spectral power across a range of acoustic frequencies (200–7000 Hz, bottom left) and output time-varying neural signals. To assess the prediction accuracy, the predicted neural signal (light lines) is compared to the original neural signal (dark lines).

Encoding Model

The neural encoding model, based on the spectrotemporal receptive field (STRF) (Theunissen et al. 2000) describes the linear mapping between the music stimulus and the high gamma neural response at individual electrodes. The encoding model was estimated as follows:

$$\hat{R}(t,\ n) = \sum_{\tau}\sum_{f} h(\tau, f, n)S(t-\tau, f),$$

where $\hat{R}(t,\ n)$ is the predicted high gamma neural activity at time t and electrode n, $S(t-\tau, f)$ is the spectrogram representation at time $(t-\tau)$, and acoustic frequency f. Finally, $h(\tau, f, n)$ is the linear transformation matrix that depends on the time lag τ, the frequency f, and electrodes n. h represents the spectrotemporal receptive field of each electrode. The neural tuning properties of a variety of stimulus parameters in different sensory systems have been assessed using STRFs (Wu et al. 2006). We used Ridge regression to fit the encoding model (Thirion et al. 2011), and a 10-fold cross-validation resampling procedure, with no overlap between training and test partitions within each resample. We performed grid search on the training set to define the penalty coefficient α and the learning rate η, using a nested loop cross-validation approach. Statistical significance of individual parameters was assessed by the z-test (mean coefficient divided by standard error of the mean across resamples). For display in the figures, model parameters with $z < 3.1$ ($P < 0.001$) were set to zero in order to emphasize only significant weights.

Decoding Model

The decoding model linearly mapped the neural activity to the music representation, as a weighted sum of activity at each electrode, as follows:

$$\hat{S}(t,\ f) = \sum_{\tau}\sum_{n} g(\tau,\ f,\ n)R(t-\tau, n),$$

where $\hat{S}(t,\ f)$ is the predicted music representation at time t and frequency f. $R(t-\tau, n)$ is the HG neural response of electrode n at time $(t-\tau)$, τ is the time lag ranging between −100 and 400 ms. Finally, $g(\tau, f, n)$ is the linear transformation matrix that depends on the time lag τ, frequency f, and electrode n. Both, neural response and music representation were synchronized, downsampled to 100 Hz, and standardized to zero mean and unit standard deviation prior to model fitting. To fit model parameters, we used gradient descent with early stopping regularization. We used a 10-fold cross-validation resampling scheme, and 20% of the training data were used as validation set to determine the early stopping criterion. Finally, model prediction accuracy was evaluated on the independent testing set, and the parameter estimates were standardized to yield the final model.

Evaluation

Prediction accuracy was quantified using the correlation coefficient (Pearson's r) between the predicted and actual HG signal using data from the independent test. Overall prediction accuracy was reported as the mean correlation over folds. The z-test was applied for all reported mean r values. Electrodes were defined as significant if the P-value was smaller than the significance threshold of $\alpha = 0.05$ (95th-percentile; FDR correction).

To further investigate the neural encoding of spectrotemporal acoustic features during music perception and music imagery, we analyzed all the electrodes that were at least significant in one condition (unless otherwise stated). Frequency tuning curves were estimated from STRFs (not thresholded), by first setting all inhibitory weights to zero (David et al. 2007), then averaging across the time dimension and converting to standardized z-scores. Frequency tuning peaks were identified as significant peak parameters in the acoustic frequency tuning curves ($z > 3.1$; $P < 0.001$)—separated by more than one-third an octave.

Decoding accuracy was assessed by calculating the correlation coefficient (Pearson's r) between the reconstructed and original music spectrogram representation using testing set data. Overall reconstruction accuracy was computed by averaging over acoustic frequencies and resamples, and standard error of the mean (SEM) was computed by taking the standard deviation across resamples. Finally, to further assess the reconstruction accuracy, we evaluated the ability to identify isolated piano notes from the test set auditory spectrogram reconstructions—using similar approach as in (Pasley et al. 2012; Martin et al. 2014).

Control Analysis for Motor Confounds

In this study, the participant played piano in 2 different conditions (music perception and imagery). Movements related to face, arm, hand, or fingers during the piano task present potential confounds to the encoding and decoding models. We controlled for possible motor confounds in 4 different ways. First, we investigated differences across conditions, as they cannot be explained by motor confounds, because movements were similar in both tasks. Second, we defined auditory sensory areas, by building encoding models on data recorded while the participant listened passively to auditory stimuli during ~35 min (speech sentences from the TIMIT corpus (Garofolo 1993) and classical music pieces by Johann Sebastian Bach—"The Art of Fugue" in C minor (BWV 1080), Contrapunctus 1–3 played by Grigory Sokolov), using the same procedure described in the section "Encoding model". Third, we built an encoding model to predict the amount of variance accounted by motor movements. For this, we built encoding models using motor-related keypresses as an input feature (keypress control condition). Keypresses were detected from the audio waveform using the MIRtoolbox (Lartillot et al. 2008). Values in the input feature were set to 1 if a keypress onset was detected or 0 if no keypress was detected. Fourth, we built additional decoding models explicitly designed to test motor versus auditory contributions: 1) a decoding model using only temporal lobe electrodes traditionally not defined as motor areas (Langers and van Dijk 2012) and 2) only auditory-responsive electrodes from the passive listening conditions (speech and music) described in the section "Decoding model".

Results

High Gamma Neural Encoding During Auditory Perception and Imagery

The participant performed 2 famous music pieces: Chopin's Prelude in C minor, Op. 28, No.20 (length of the piece played in the perception condition = 122.3 s and in the imagery condition = 120.2 s) and Bach's Prelude in C major BWV 846 (length of the piece played in the perception condition = 112.4 s and in the imagery condition = 111.2 s). Example auditory spectrograms from Chopin's Prelude determined through the participant's keypresses with the electronic piano are shown in Figure 2*B* for both perception and imagery conditions. To compare spectrotemporal auditory representations during music perception and music imagery tasks, we fit separate spectrotemporal receptive field

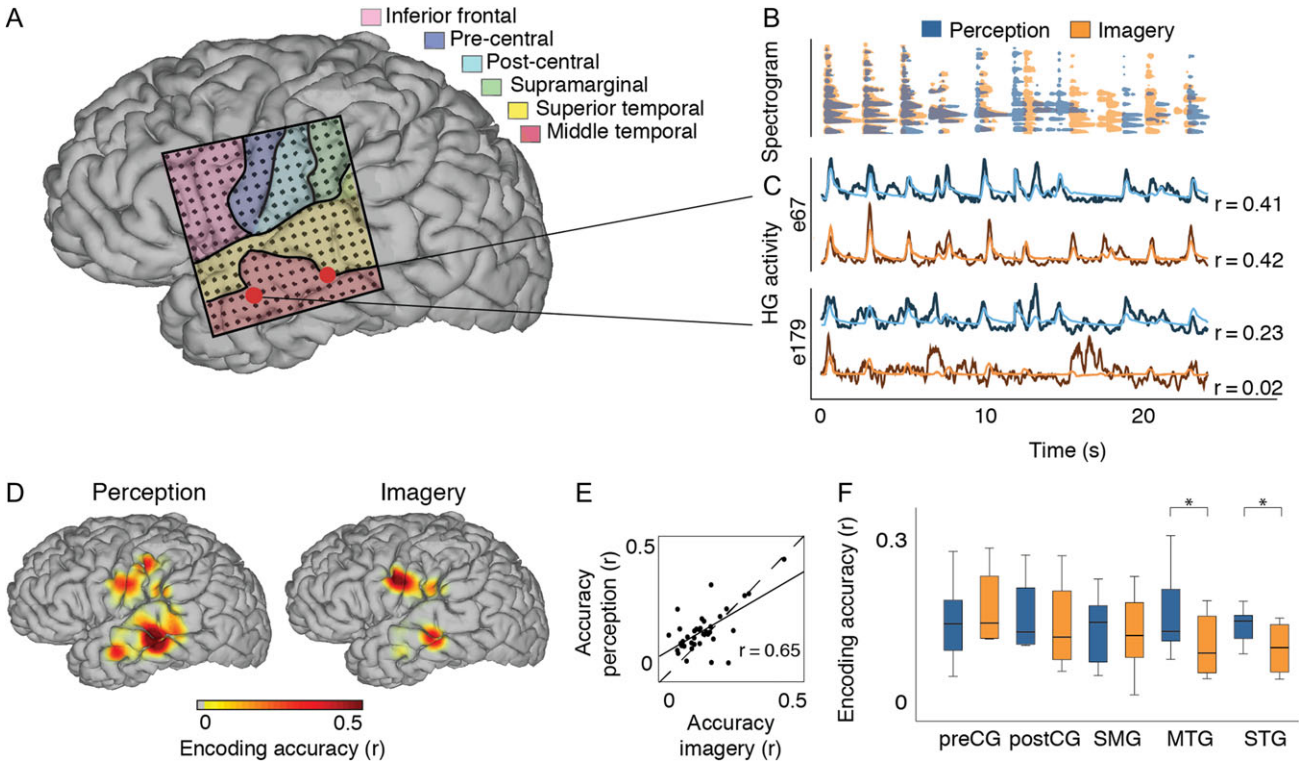

Figure 2. Prediction accuracy. (*A*) Electrode location overlaid on cortical surface reconstruction of the participant's cerebrum. (*B*) Overlay of the spectrogram contours for the perception (blue) and imagery (orange) condition (10% of maximum energy from the spectrograms) corresponding to a segment of Chopin's prelude. (*C*) Actual and predicted high gamma band power (70–150 Hz) induced by the music perception and imagery segment in (*B*). Electrode 67 has very similar predictive power across conditions, whereas electrode 179 has significantly better predictive power for perception compared to imagery. Recordings are from 2 different temporal lobe sites highlighted in pink in (*A*). (*D*) Prediction accuracy is plotted on the cortical surface reconstruction of the participant's cerebrum (map thresholded at $P < 0.05$; FDR correction). (*E*) Prediction accuracy of significant electrodes of the perception model as a function of the imagery model. Electrode-specific prediction accuracy is correlated between perception and imagery models ($r = 0.65$; $P < 10^{-4}$; randomization test). (*F*) Prediction accuracy as a function of anatomic location (pre-central gyrus (pre-CG), post-central gyrus (post-CG), supramarginal gyrus (SMG), medial temporal gyrus (MTG), and superior temporal gyrus (STG)).

(STRF) models in each condition. We used these models to quantify specific anatomical and neural tuning differences between auditory perception and imagery.

For both perception and imagery conditions, the observed and predicted high gamma neural responses are illustrated for 2 individual electrodes in the temporal lobe, respectively (Fig. 2*C*), together with the corresponding music spectrum (Fig. 2*B*). The predicted neural response for electrode 67 was correlated with its corresponding measured neural response in both perception ($r = 0.41$; $P < 10^{-7}$; one-sample z-test; FDR correction) and imagery ($r = 0.42$; $P < 10^{-4}$; one-sample z-test; FDR correction) conditions. The predicted neural response for electrode 179 was correlated with the actual neural response only in the perception condition ($r = 0.23$; $P < 0.005$; one-sample z-test; FDR correction) but not in the imagery condition ($r = -0.02$; $P > 0.5$; one-sample z-test; FDR correction). The difference between both conditions was significant for electrode 179 ($P < 0.05$; 2-sample *t*-test), but not for electrode 67 ($P > 0.5$; 2-sample *t*-test). This suggests that there is an underlying relationship between time-varying imagined sound features and STG neural activity, but that this relationship is dependent on cortical location.

To further investigate anatomical similarities and differences between the perception and imagery conditions, we plotted the anatomical layout of prediction accuracy of individual electrodes. In both conditions, results showed that sites with the highest prediction accuracy in both conditions were located in the superior and middle temporal gyrus, pre- and post-central gyrus, and supramarginal gyrus (Fig. 2*D*; heat map thresholded to $P < 0.05$; one-sample z-test; FDR correction). These results were overlapping with auditory areas (Fig. 7*B* and S3; see Material and Methods for details), and consistent with previous findings showing the presence of STRFs in the temporal lobe (STG, MTG), as well as sensorimotor cortex (pre- and post-central gyrus, supramarginal gyrus) (Pasley et al. 2012; Mesgarani et al. 2014; Cheung et al. 2016).

Among the 256 electrodes recorded, 210 were used in the STRF analysis, while the remaining 46 electrodes were removed due to excessive noise (epileptic artifacts, broadband electromagnetic noise from hospital equipment, or poor contact with the cortical surface). Within the analyzed electrodes, 35 and 15 electrodes had significant prediction accuracy in the perception and imagery condition, respectively ($P < 0.05$; one-sample z-test; FDR correction), while 9 electrodes were significant in both conditions (Fig. S1). To compare the prediction accuracy across conditions, we performed additional analysis on the electrodes that had significant accuracy in at least one condition (41 electrodes; unless otherwise stated). Prediction accuracy of individual electrodes was correlated between perception and imagery (Fig. 2*E*; 41 electrodes; $r = 0.65$; $P < 10^{-4}$; randomization test). Because both perception and imagery STRF models are based on the same auditory stimulus representation, the correlated prediction accuracy provides strong evidence for a shared neural representation of sound based on spectrotemporal features.

To assess how brain areas encoding auditory features varied across experimental conditions, we analyzed the significant electrodes in the gyri highlighted in Figure 2A (pre-central gyrus (pre-CG), post-central gyrus (post-CG), supramarginal gyrus (SMG), medial temporal gyrus (MTG), and superior temporal gyrus (STG)) using Wilcoxon signed-rank test ($P > 0.05$; one-sample Kolmogorov–Smirnov test; Fig. 2F). Results showed that the encoding accuracy in the MTG and STG was higher for the perception (MTG: $M = 0.16$, STG: $M = 0.13$) than for the imagery (MTG: $M = 0.11$, STG: $M = 0.08$; $P < 0.05$; Wilcoxon signed-rank test; Bonferroni correction). The encoding accuracy in the pre-CG, post-CG and SMG was not different between the perception (pre-CG $M = 0.17$; post-CG $M = 0.15$; SMG $M = 0.12$; $P > 0.5$; Wilcoxon signed-rank test; Bonferroni correction) and imagery (pre-CG $M = 0.14$; post-CG $M = 0.13$; SMG $M = 0.12$; $P > 0.5$; Wilcoxon signed-rank test; Bonferroni correction) conditions. The significant improvement of the perception versus imagery model was thus specific to the temporal lobe, which may reflect underlying differences in spectrotemporal encoding mechanisms, or alternatively, a greater sensitivity to discrepancies between the actual content of imagery and the recorded sound stimulus used in the model.

Spectrotemporal Tuning During Auditory Perception and Imagery

To quantify similarities and differences in neural tuning, we used the STRF models described above which were fit separately for perception versus imagery experimental conditions to allow comparison of neural tuning between the 2 conditions. Examples of perception and imagery STRFs are shown in Figure 3A for temporal electrodes (Fig. S2 for STRFs at all electrodes). These STRFs highlight neural stimulus preferences as shown by the excitatory (warm color) and inhibitory (cold color) subregions. The similarities (correlation coefficients) between the vectorized STRFs in the perception and imagery condition are plotted on the surface reconstruction of the participant's brain for electrodes that had significant prediction accuracy in at least one condition. Correlations coefficients ranged between 0.1 and 0.8. (Fig. 3B). STRF similarity was significantly correlated with prediction accuracy similarity, as defined by the relative change ($\text{abs}((\text{acc}_{\text{perception}} - \text{acc}_{\text{imagery}})/(\text{acc}_{\text{perception}} + \text{acc}_{\text{imagery}}))$) across electrodes ($r = -0.40$, $P < 0.05$), suggesting that electrodes which predicted well across both conditions also had similar spectrotemporal tuning in the 2 conditions. Overall similarities in STRF tuning and prediction accuracy suggest a shared auditory representation between auditory imagery and perception.

In addition to the overall STRF structure, we analyzed similarity in temporal and frequency tuning independently. Figure 4A shows the correlation in latencies for the perception and imagery conditions, defined as the time lag of the maximum deviation in the standardized STRF for electrodes that are significant in at least one condition. The peak latency correlated between conditions ($r = 0.43$; $P < 0.005$; randomization test), and the mean peak latency between conditions was not significantly different (mean perception = 79 ms and mean imagery = 82 ms; $P > 0.5$; 2-sample t-test; $P > 0.5$; 2-sample t-test). This suggests that neural activity evoked by the perceptual and imagery processes had similar temporal offsets relative the piano keypress. We next analyzed frequency tuning curves estimated from the STRFs (see Materials and Methods for details). Examples of frequency tuning curves for both perception and imagery encoding models are shown for the electrodes indicated by the black outline in the anatomic brain (Fig. 4B). Across conditions, the majority of individual electrodes exhibited a complex frequency tuning profile. For each electrode, similarities between the frequency tuning curves in the perception and imagery models were quantified using Pearson's correlation coefficient. The anatomical distribution of frequency tuning curve similarity is plotted in Figure 4B, with the correlation at individual sites ranging between $r = -0.3$ and 0.6. Different electrodes are sensitive to different acoustic frequencies important for auditory processing. We next assessed how frequency tuning of predictive electrodes ($N = 41$) varied during the 2 conditions. First, to evaluate how the acoustic spectrum was covered at the population level, we quantified the proportion of significant electrodes with a tuning peak at each acoustic frequency. Figure 4C depicts the number of electrodes with significant STRF tuning for each frequency bin for the perception (blue) and imagery (orange) conditions. Note that the maximum fundamental frequency of the keyboard was ~4200 Hz but acoustic energy is present in higher frequencies due to harmonics (Fig. 5B). Tuning peaks were identified as significant parameters in the acoustic frequency tuning curves ($z > 3.1$; $P < 0.001$;

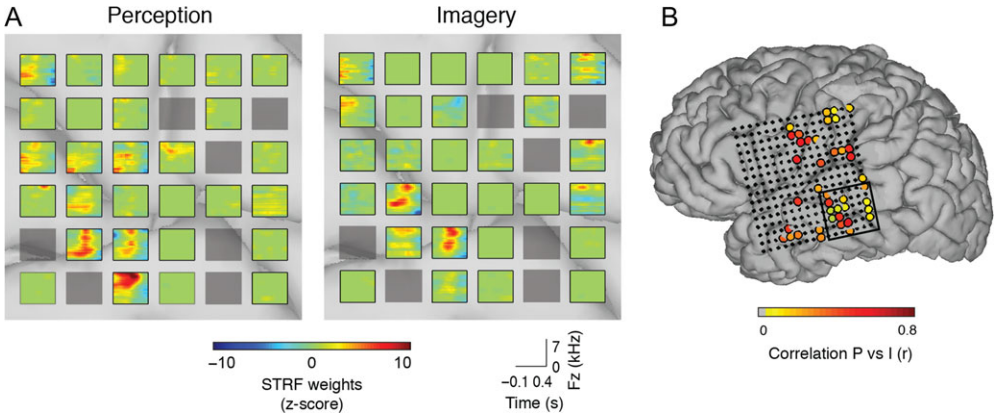

Figure 3. Spectrotemporal receptive fields. (A) Examples of standard STRFs for the perception (left panel) and imagery (right panel) models (warm colors indicate where the neuronal ensemble is excited, cold colors indicate where the neuronal ensemble is inhibited). Electrodes whose STRFs are shown are outlined in black in (B). Gray electrodes were removed from the analysis due to excessive noise (see Materials and Methods). (B) The correlation coefficients between the vectorized STRFs in the perception and imagery condition are plotted on the surface reconstruction of the participant's brain for electrodes that had significant prediction accuracy in at least one of the perception and imagery conditions.

separated by more than one-third an octave). The proportion of electrodes with tuning peaks was larger for the perception (mean = 0.19) than for the imagery (mean = 0.14) condition (Fig. 4C; $P < 0.05$; Wilcoxon signed-rank test). Results showed that some sites exhibit significant frequency tuning at frequencies <500 Hz and >6000 Hz during the imagery condition. This indicates that neural activity at these sites systematically increased with the (assumed ground truth) imagery contents of high and low frequencies. If the pianist were unable to imagine this frequency range, we would expect nonsignificant tuning (weights close to 0) at high or low frequencies because there would be no systematic, time-locked relationship between an increase in neural activity and the low or high frequencies produced by the keypresses during imagery. Over the full range of the acoustic frequency spectrum, both conditions exhibited reliable frequency selectivity. The fraction of acoustic frequency bins covered with peaks by predictive electrodes was 0.91 for the perception and 0.88 for the imagery. These findings showed robust spectrotemporal receptive fields during auditory imagery with substantial, but not complete overlap in frequency tuning and cortical location compared to receptive fields measured during auditory perception. This is in accordance with previous research that showed partial overlap across conditions (Griffiths 1999; Kosslyn et al. 2001; Yoo et al. 2001; Halpern et al. 2004; Bunzeck et al. 2005).

Reconstruction of Auditory Features During Music Perception and Imagery

An additional indication showing that auditory perceptual elements of sounds are represented in the brain is to reconstruct the auditory spectrogram representations from high gamma neural signals. Results showed that the overall reconstruction accuracy was higher than zero in both conditions (Fig. 5A; left panel; $P < 0.001$; randomization test), but did not differ between conditions ($P > 0.05$; 2-sample t-test). As a function of acoustic frequency, mean accuracy ranged from $r = 0$ to 0.45 (Fig. 5A; right panel). These results showed for the first time that acoustic features can be accurately decoded from subjective experience of music imagery.

We further assessed reconstruction accuracy by evaluating the ability to identify isolated piano notes from the test set auditory spectrogram reconstructions. Examples of original and reconstructed segments are depicted in Figure 5B for the perception (left) and imagery model (right). For the identification, we extracted 0.5-s segments at piano note onsets from the original and reconstructed auditory spectrogram. Note that onsets were defined as the maxima of the onset detection curve (amplitude envelop of the spectrogram), using the MIRtoolbox (Lartillot et al. 2008). Then, we computed the correlation coefficient between a target reconstructed spectrogram and original spectrograms in

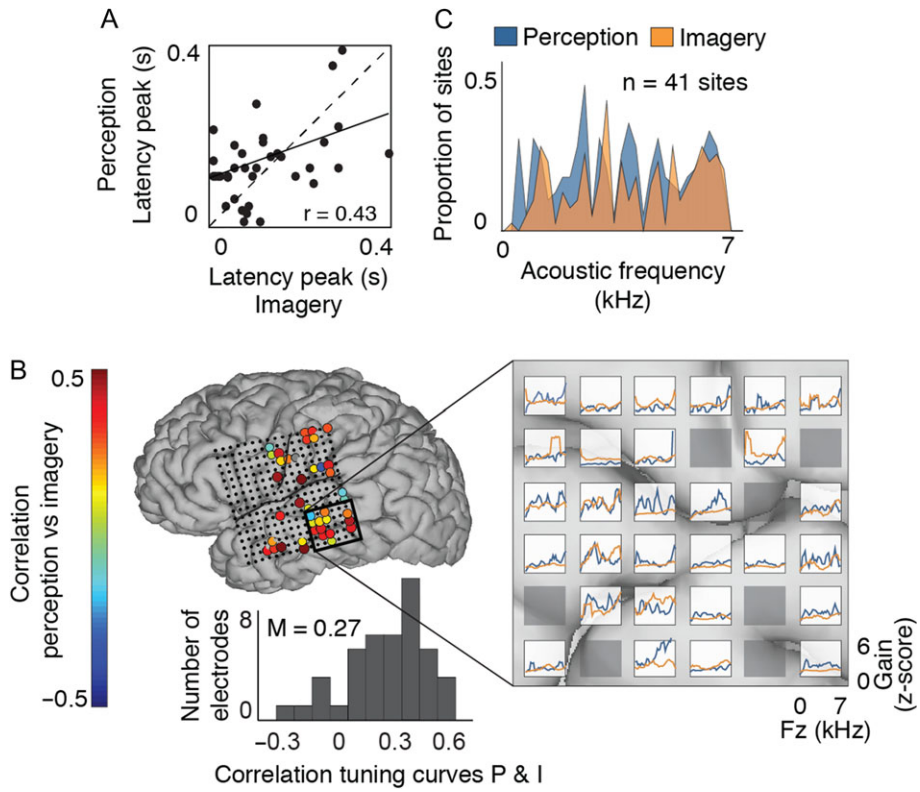

Figure 4. Auditory tuning. (*A*) Peak latency estimated from STRFs was significantly correlated between perception and imagery conditions ($r = 0.43$; $P < 0.005$; randomization test). (*B*) Examples of frequency tuning curves (right) for perception and imagery encoding models (averaged over the time lag dimension of the STRF). Black outline in the surface reconstruction of the patient's brain (left) indicates electrode location. Gray electrodes were removed from the analysis due to excessive noise. Correlation coefficients between the perception and imagery frequency tuning curves are plotted for significant electrodes on the cortical surface reconstruction (left). The bottom panel plots the histogram of electrode correlation coefficients between perception and imagery frequency tuning. (*C*) Proportion of predictive electrode sites ($N = 41$) with peak tuning at each frequency. Tuning peaks were identified as significant parameters in the acoustic frequency tuning curves ($z > 3.1$; $P < 0.001$) and separated by more than one-third of an octave.

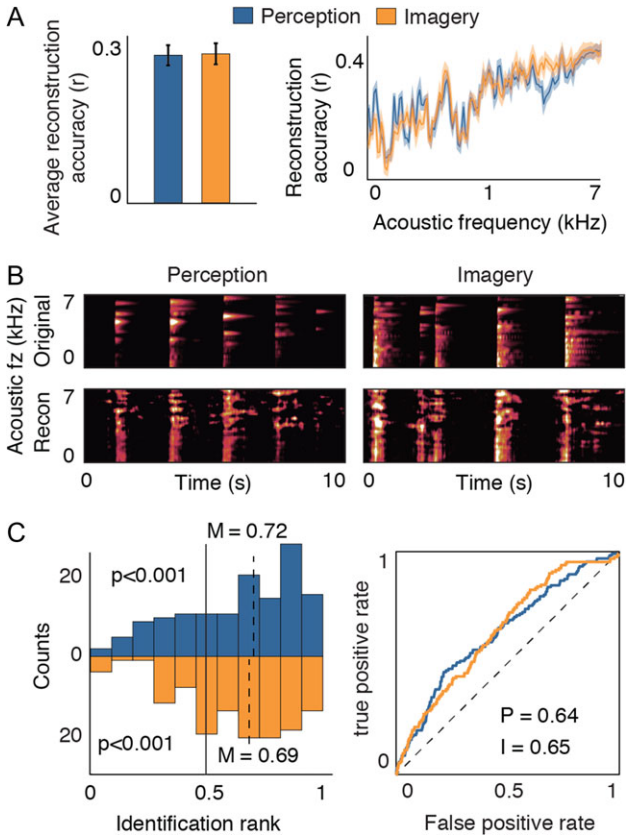

Figure 5. Reconstruction accuracy. (*A*) Left panel, overall reconstruction accuracy of the spectrogram representation for perception (blue) and imagery (orange) conditions. Error bars denote SEM. Right panel, reconstruction accuracy as a function of acoustic frequency. Shaded region denotes SEM. (*B*) Examples of original and reconstructed segments for the perception (left) and the imagery (right) model. (*C*) Left panel, distribution of identification rank for all reconstructed spectrogram ($N = 140$ for perception and $N = 135$ for imagery). Median identification rank is 0.65 and 0.63 for the perception and imagery decoding model, respectively, which is significantly higher than 0.50 chance level ($P < 0.001$; randomization test). Right panel, receiver operating characteristic (ROC) plot of identification performance for the perception (blue curve) and imagery (orange curve) model. Diagonal black line indicates no predictive power.

the candidate set. Finally, we sorted the coefficients and computed the identification rank as the percentile rank of the correct spectrogram. This metric reflects how well the target reconstruction matched the correct original spectrogram out of all candidate. Results showed that the median identification rank of individual piano notes was significantly higher than chance for both conditions (Fig. 5C; left panel; median identification rank perception = 0.72 and imagery = 0.69; $P < 0.001$; randomization test). Similarly, the area under the curve (AUC) of identification performance for the perception (blue curve) and imagery (orange curve) model was well above chance level (Fig. 5C; right panel; diagonal black dashed line indicates no predictive power; $P < 0.001$; randomization test).

Cross-condition Analysis

Another way to evaluate the overlapping degree between both perception and imagery conditions is to apply the decoding model built in the perception condition to imagery neural data, and vice-versa. This approach is based on the hypothesis that both tasks share neural mechanisms and is useful when one of

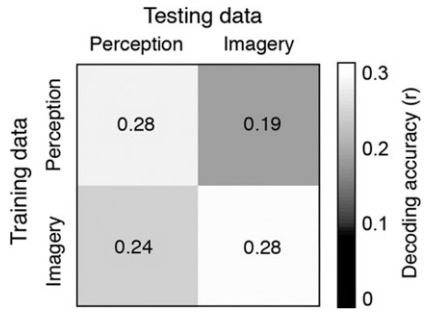

Figure 6. Cross-condition analysis. Reconstruction accuracy when the decoding model was trained on the perception condition and applied to the imagery neural data and vice-versa. Decoding performance improved by 50% when the model was trained and tested on the imagery condition ($r = 0.28$; $P < 0.001$; randomization test), compared to when the perception model was applied to imagery data ($r = 0.19$; $P < 0.001$; randomization test).

the models cannot be built directly, because of the lack of observable measures. This technique has been successfully applied to various fields, such as vision (Haynes and Rees 2005; Reddy et al. 2010; Horikawa et al. 2013) and speech (Martin et al. 2014). When the model was trained on the imagined condition and tested on the imagined condition ($r = 0.28$; $P < 0.001$; randomization test), decoding performance significantly improved by 50% ($P < 0.0001$; one-sided Hoteling's *t*-test) compared to when the perception model was applied to imagined data ($r = 0.19$; $P < 0.001$; randomization test; Fig. 6). This highlight the importance of having a model that is specific to each condition, and also emphasize that these results are not based on movements per se, as these are equal across conditions, thus should give equal detection results.

Control Analysis for Motor Confounds

We controlled for possible motor confounds in 4 different ways. First, differences across conditions cannot be explained by motor confounds, because movements were similar in both tasks. For instance, the STRFs of one temporal lobe electrode (electrode 67) are correlated between perception and imagery conditions (Fig. 7A; $r = 0.76$; $P < 0.05$; Bonferroni correction) and the encoding accuracies at this electrode are similar across conditions (mean perception = 0.41; mean imagery = 0.42; $P > 0.05$; 2-sample *t*-test). The STRFs in an adjacent electrode (electrode 68) exhibit distinct, uncorrelated tuning patterns when measured during the 2 conditions (Fig. 7A; $r = 0.04$; $P > 0.05$; Bonferroni correction) and different encoding accuracies (mean perception = 0.15; mean imagery = 0.03; $P < 0.05$; 2-sample *t*-test). Because essentially the same motor sequence was present in both the perception and imagery conditions, the effects of this motor sequence at each individual site would be expected to be similar across conditions. This electrode-specific control for motor sequence suggests that motor activity is unlikely to explain the differences in neural tuning and prediction accuracies observed between perception and imagery. Second, brain areas that significantly encoded music perception and imagery overlapped with auditory sensory areas (Fig. 7B and Fig. S3), as revealed by the prediction accuracy and STRFs during passive listening (no movement) to TIMIT sentences and classical music. Although motor commands have been shown to modulate auditory responses (Zatorre et al. 2007), it is unlikely that the motor commands associated with pressing piano keys can induce prediction accuracies and STRFs that are

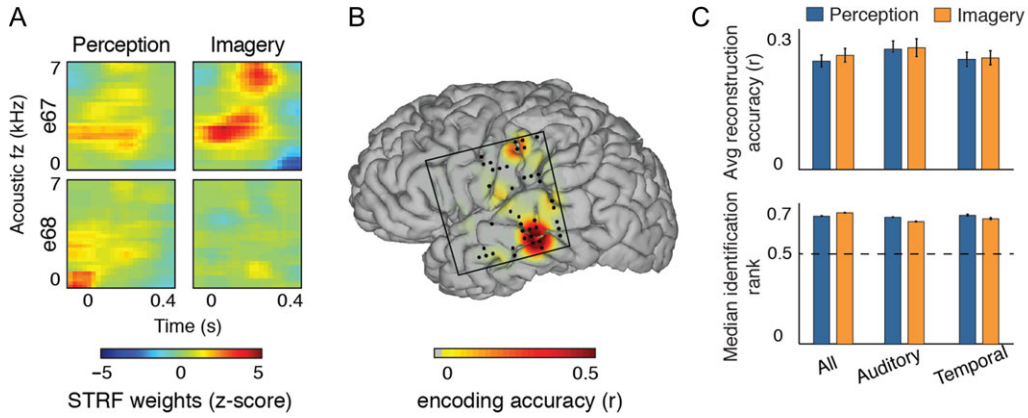

Figure 7. Control analysis for motor confound. (A) STRFs for 2 neighboring electrodes for perception (left) and imagery (right) encoding models. For electrode 67, the STRF is strongly correlated between perception and imagery conditions ($r = 0.04$), while there is a nonsignificant correlation ($r = 0.76$) in the adjacent electrode 68. (B) Prediction accuracy plotted on the cortical surface reconstruction of the participant's brain (map thresholded at $P < 0.05$; FDR correction) for the passive listening data sets (speech and music). Black dots represent electrodes that had significant prediction accuracy in at least one of the perception and imagery conditions. (C) Overall reconstruction accuracy (upper panel) and median identification rank (lower panel) when using all electrodes, only temporal electrodes, or only auditory-response electrodes (see Materials and methods for details).

overlapping with those obtained with pure auditory listening. Third, we evaluated a control encoding model based on motor patterns alone to quantify the prediction accuracy that could be accounted for by movements associated with the keypress (see Materials and Methods for details). The encoding prediction accuracies for the keypress control condition were not significant ($P > 0.05$; one-sample z-test; FDR correction). Finally, we built 2 additional decoding models, using 1) only temporal lobe electrodes and 2) only auditory-responsive electrodes (Fig. 7C; see Materials and Methods for details). Both models showed significant reconstruction accuracy ($P < 0.001$; randomization test) and identification rank ($P < 0.001$; randomization test). This indicates that even using only auditory-responsive electrodes (thus removing the more motor-driven electrodes), we are still able to reconstruct the spectrograms significantly above chance levels.

Discussion

Experimental studies of auditory imagery are difficult due to the subjective nature and absence of verifiable and observable measures. The task design in the current study allowed precise time-locking between the recorded neural activity and spectrotemporal features of music imagery, alleviating a core methodological issue in most imagery research. This approach provided a unique opportunity to quantitatively study neural encoding during auditory imagery and compare neural tuning properties with auditory perception. We describe the first evidence of spectrotemporal receptive fields and neural tuning to auditory features during music imagery and provide quantitative comparison to the neural encoding of actual music perception in the same cortical sites. In particular, we observed that neuronal ensembles were tuned to acoustic frequencies during imagined music, suggesting that neural encoding of spectral features occurs in the absence of actual perceived sound. This is in agreement with previous studies showing increased neural activity during sound imagery (Zatorre et al. 1996; Griffiths 1999; Halpern and Zatorre 1999; Rauschecker 2001; Halpern et al. 2004; Kraemer et al. 2005), as well as studies of auditory neural encoding during "restored speech", when a speech instance is replaced by noise, but the listener instead perceives a specific speech sound (Holdgraf et al. 2016; Leonard et al. 2016).

Importantly, the current results showed substantial, but not complete overlap in functional and anatomical patterns of neural encoding for music perception compared to imagery. Across the population of ECoG electrodes, spectral and temporal tuning properties, as well as anatomical location, showed strong correlations between perception and imagery, yet specific electrodes revealed significantly different spectrotemporal tuning between the 2 conditions. Similar patterns of partial functional and anatomical overlap have been found in the visual system (Kosslyn and Thompson 2000, 2003) and in the motor system (Roth et al. 1996; Miller et al. 2010). Brain areas with significant prediction accuracy were located in the superior and middle temporal gyrus, pre- and post-central gyrus, and supramarginal gyrus, and overlapped with auditory-responsive regions determined from separate passive listening data sets. STRFs outside of traditional auditory cortex, in pre- and post-central gyrus, and supramarginal gyrus, have been observed in a number of ECoG studies (Pasley et al. 2012; Martin et al. 2014; Mesgarani et al. 2014; Cheung et al. 2016), and the functional role of these sites outside of temporal cortex remains an open research question. These findings are in agreement with neuroimaging studies that have consistently reported activity in motor-related areas when imagining the sound of musical excerpts (Zatorre and Halpern 2005) or imagining performing instruments (Langheim 2002; Meister et al. 2004).

Our results also showed that auditory features can be reconstructed from neural activity of the imagined music and used to identify isolated piano notes from the reconstructed auditory spectrograms. Because both perception and imagery models are based on the same auditory stimulus representation, the correlated prediction accuracy provides strong evidence for a shared neural representation of sound based on spectrotemporal features, as suggested by previous behavioral and brain lesion studies (see Hubbard 2010 for a review). These results build on earlier studies showing anatomical and behavioral similarities between perception and imagery (Griffiths 1999; Yoo et al. 2001; Halpern et al. 2004; Bunzeck et al. 2005).

An important advantage of the encoding model approach is that it describes not only the anatomical location of imagery neural processes (as with prior neuroimaging work: Griffiths 1999; Kraemer et al. 2005; Zatorre et al. 2009), but also how neural tuning to specific stimulus features is organized during

auditory imagery. Our results provide an explicit characterization, beyond anatomical overlap, of the underlying neural representations of auditory perception and imagery. One limitation of the encoding model approach is the possibility that alternative models based on correlated stimulus representations may be a more accurate description of the neural response. For example, a nonlinear sound representation such as the modulation power spectrum (Chi et al. 1999) is correlated with the spectrotemporal representation studied here and may yield higher prediction accuracy because it models additional nonlinear neural processes not accounted for by a spectrogram-based representation (Pasley et al. 2012). Direct comparison of such alternative encoding models to the STRF is an interesting avenue for future work and may identify auditory features that provide a more accurate description of the underlying imagery neural representation.

This study involved active motor movements associated with the piano task and could represent a potential experimental confound. Sensory-motor interactions during musical performance are well-known, and identification of auditory versus motor processes remains challenging (see Zatorre et al. 2007 for a review). For example, it is possible that practicing a musical piece over several months could create a mapping where the motor sequence itself induces tonotopic auditory responses in the associated auditory cortex. To address possible motor confounds, we used 4 distinct analyses. First, we observed differences across conditions, which cannot be explained by motor-related neural activity, given that the same movement sequences were performed in both conditions. Second, we found that brain areas involved during the perception and imagery conditions overlapped with auditory areas, suggesting that these were auditory sensory brain responses rather than movement related neural activity. Third, the keypress control model did not account for variability in the neural responses, suggesting that the brain activity was not correlated with hand movement as defined by keypresses. Finally, we were able to decode spectrotemporal features and identify piano keys using models built only with auditory-responsive electrodes.

Studies have shown the importance of both hemispheres for auditory perception and imagination (Zatorre and Halpern 1993; Zatorre et al. 1996; Griffiths 1999; Rauschecker 2001; Halpern et al. 2004; Kraemer et al. 2005). In our task, the grid was located on the left hemisphere, and allowed significant encoding and decoding accuracy within high gamma frequency ranges, consistent with the notion that music perception and imagery processes are also evident in the left hemisphere.

Methodological issues in investigating imagery are numerous, including the lack of evidence that the desired mental task was operational. The current task design did not allow verifying how the mental task was performed, although the behavioral index of keypress on the piano was utilized to indicate the precise time and frequency content of the intended imagined sound. In addition, higher cognitive functions, such as attention and other task-related processes can rapidly alter neural selectivity and STRF structure (Fritz et al. 2003; Atiani et al. 2009; Mesgarani et al. 2009; David et al. 2012; Ding and Simon 2012; Mesgarani and Chang 2012). While the current study focused on music imagery, neural selectivity might be altered during imagery of other auditory stimuli, such as speech or environmental sounds, if different attentional or task-related mechanisms are engaged. Finally, we recorded a skilled piano player, and it has been suggested that participants with musical training exhibited better pitch and temporal acuity during auditory imagery than did participants with little or no musical training (Janata and Paroo 2006; Herholz et al. 2008). Furthermore, tonotopic maps located in the STG are enlarged within trained musicians (Pantev et al. 1998). Thus, having a trained piano player may have contributed to improved auditory imagery ability (see also Halpern 1988; Zatorre and Halpern 1993; Zatorre et al. 1996), and reduced issues related to spectral and temporal errors. Given the music proficiency of this single participant, our results might not be representative of the general population and further investigations are needed to assess with participants with less musical background.

Supplementary Material

Supplementary data is available at *Cerebral Cortex* online.

Funding

This research was supported by NINDS Grant R3721135, DARPA D16PC00053, the Zeno-Karl Schindler Foundation, Kavli Institute for Brain and Mind Innovative Research grant, NIH grants F32-DC013486, 5K99DC01 2804, R00-NS065120, DP2-OD00862, R01-DC012379, R01DC5779, European Research Council Advanced Grant. The New York Stem Cell Foundation, The McKnight Foundation, The Shurl and Kay Curci Foundation, The William K Bowes Foundation and The Nielsen Corporation.

Notes

Conflict of Interest: None declared.

References

Aertsen AMHJ, Olders JHJ, Johannesma PIM. 1981. Spectro-temporal receptive fields of auditory neurons in the grass-frog: III. Analysis of the stimulus-event relation for natural stimuli. Biol Cybern. 39:195–209.

Aleman A, Nieuwenstein MR, Böcker KB, de Haan EH. 2000. Music training and mental imagery ability. Neuropsychologia. 38:1664–1668.

Atiani S, Elhilali M, David SV, Fritz JB, Shamma SA. 2009. Task difficulty and performance induce diverse adaptive patterns in gain and shape of primary auditory cortical receptive fields. Neuron. 61:467–480.

Boonstra TW, Houweling S, Muskulus M. 2009. Does asynchronous neuronal activity average out on a macroscopic scale? J Neurosci. 29:8871–8874.

Brodsky W, Henik A, Rubinstein B-S, Zorman M. 2003. Auditory imagery from musical notation in expert musicians. Percept Psychophys. 65:602–612.

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

Cheung C, Hamiton LS, Johnson K, Chang EF. 2016. The auditory representation of speech sounds in human motor cortex. eLife. 5:12577.

Chi T, Gao Y, Guyton MC, Ru P, Shamma S. 1999. Spectro-temporal modulation transfer functions and speech intelligibility. J Acoust Soc Am. 106:2719–2732.

Chi T, Ru P, Shamma SA. 2005. Multiresolution spectrotemporal analysis of complex sounds. J Acoust Soc Am. 118:887.

Clopton BM, Backoff PM. 1991. Spectrotemporal receptive fields of neurons in cochlear nucleus of guinea pig. Hear Res. 52:329–344.

Crone NE, Boatman D, Gordon B, Hao L. 2001. Induced electrocorticographic gamma activity during auditory perception. Clin Neurophysiol. 112:565–582.

David SV, Fritz JB, Shamma SA. 2012. Task reward structure shapes rapid receptive field plasticity in auditory cortex. Proc Natl Acad Sci USA. 109:2144–2149.

David SV, Mesgarani N, Shamma SA. 2007. Estimating sparse spectro-temporal receptive fields with natural stimuli. Netw Bristol Engl. 18:191–212.

Ding N, Simon JZ. 2012. Emergence of neural encoding of auditory objects while listening to competing speakers. Proc Natl Acad Sci. 109:11854–11859.

Eggermont JJ, Aertsen AMHJ, Johannesma PIM. 1983. Quantitative characterisation procedure for auditory neurons based on the spectro-temporal receptive field. Hear Res. 10:167–190.

Fritz J, Shamma S, Elhilali M, Klein D. 2003. Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. Nat Neurosci. 6:1216–1223.

Garofolo JS 1993. TIMIT acoustic-phonetic continuous speech corpus.

Griffiths TD. 1999. Human complex sound analysis. Clin Sci. 96: 231–234.

Halpern AR. 1988. Mental scanning in auditory imagery for songs. J Exp Psychol Learn Mem Cogn. 14:434–443.

Halpern AR. 1989. Memory for the absolute pitch of familiar songs. Mem Cognit. 17:572–581.

Halpern AR. 2001. Cerebral substrates of musical imagery. Ann N Y Acad Sci. 930:179–192.

Halpern AR. 2015. Differences in auditory imagery self-report predict neural and behavioral outcomes. Psychomusicol Music Mind 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.

Haynes J-D, Rees G. 2005. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. Nat Neurosci. 8:686–691.

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 B, Humphries C, Muftuler T. 2003. Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt. J Cogn Neurosci. 15: 673–682.

Holdgraf CR, de Heer W, Pasley B, Rieger J, Crone N, Lin JJ, Knight RT, Theunissen FE. 2016. Rapid tuning shifts in human auditory cortex enhance speech intelligibility. Nat Commun. 7:13654.

Horikawa T, Tamaki M, Miyawaki Y, Kamitani Y. 2013. Neural decoding of visual imagery during sleep. Science. 340: 639–642.

Hubbard TL. 2010. Auditory imagery: empirical findings. Psychol Bull. 136:302–329.

Huth AG, de Heer WA, Griffiths TL, Theunissen FE, Gallant JL. 2016. Natural speech reveals the semantic maps that tile human cerebral cortex. Nature. 532:453–458.

Intons-Peterson M. 1992. Components of auditory imagery. In: Reisberg D, editor. Auditory imagery. Hillsdale, NJ: L. Erlbaum Associates. p. 45–72.

Janata P, Paroo K. 2006. Acuity of auditory images in pitch and time. Percept Psychophys. 68:829–844.

Kosslyn SM, Ganis G, Thompson WL. 2001. Neural foundations of imagery. Nat Rev Neurosci. 2:635–642.

Kosslyn SM, Thompson WL. 2000. Shared mechanisms in visual imagery and visual perception: Insights from cognitive neuroscience. In: Gazzaniga MS, editor. The new cognitive neurosciences. 2nd ed. Cambridge, MA: MIT Press.

Kosslyn SM, Thompson WL. 2003. When is early visual cortex activated during visual mental imagery? Psychol Bull. 129: 723–746.

Kraemer DJM, Macrae CN, Green AE, Kelley WM. 2005. Musical imagery: sound of silence activates auditory cortex. Nature. 434:158–158.

Lachaux J-P, Axmacher N, Mormann F, Halgren E, Crone NE. 2012. High-frequency neural activity and human cognition: past, present and possible future of intracranial EEG research. Prog Neurobiol. 98:279–301.

Langers DRM, van Dijk P. 2012. Mapping the tonotopic organization in human auditory cortex with minimally salient acoustic stimulation. Cereb Cortex. 22:2024–2038.

Langheim F. 2002. Cortical systems associated with covert music rehearsal. Neuroimage. 16:901–908.

Lartillot O, Toiviainen P, Eerola T. 2008. A matlab toolbox for music information retrieval. In: Preisach C, Burkhardt H, Schmidt-Thieme L, Decker R, editors. Data analysis, machine learning and applications. Berlin, Heidelberg: Springer Berlin Heidelberg. p. 261–268.

Leonard MK, Baud MO, Sjerps MJ, Chang EF. 2016. Perceptual restoration of masked speech in human cortex. Nat Commun. 7:13619.

Lima CF, Lavan N, Evans S, Agnew Z, Halpern AR, Shanmugalingam P, Meekings S, Boebinger D, Ostarek M, McGettigan C, et al. 2015. Feel the noise: relating individual differences in auditory imagery to the structure and function of sensorimotor systems. Cereb Cortex. 25:4638–4650.

Llorens A, Trébuchon A, Liégeois-Chauvel C, Alario F-X. 2011. Intra-cranial recordings of brain activity during language production. Front Psychol. 2:375.

Lotte F, Brumberg JS, Brunner P, Gunduz A, Ritaccio AL, Guan C, Schalk G. 2015. Electrocorticographic representations of segmental features in continuous speech. Front Hum Neurosci. 09:97.

Lotze M, Scheler G, Tan H-RM, Braun C, Birbaumer N 2003. The musician's brain: functional imaging of amateurs and professionals during performance and imagery. Neuroimage 20: 1817–1829.

Martin S, Brunner P, Holdgraf C, Heinze H-J, Crone NE, Rieger J, Schalk G, Knight RT, Pasley BN. 2014. Decoding spectrotemporal features of overt and covert speech from the human cortex. Front Neuroeng. 7:14.

Meister IG, Krings T, Foltys H, Boroojerdi B, Müller M, Töpper R, Thron A. 2004. Playing piano in the mind—an fMRI study on music imagery and performance in pianists. Brain Res Cogn Brain Res. 19:219–228.

Mesgarani N, Chang EF. 2012. Selective cortical representation of attended speaker in multi-talker speech perception. Nature. 485:233–236.

Mesgarani N, Cheung C, Johnson K, Chang EF. 2014. Phonetic feature encoding in human superior temporal gyrus. Science. 343:1006–1010.

Mesgarani N, David SV, Fritz JB, Shamma SA. 2009. Influence of context and behavior on stimulus reconstruction from

neural activity in primary auditory cortex. J Neurophysiol. 102:3329–3339.
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 melodies. Music Percept Interdiscip J. 12:175–197.
Miller KJ, Leuthardt EC, Schalk G, Rao RPN, Anderson NR, Moran DW, Miller JW, Ojemann JG. 2007. Spectral changes in cortical surface potentials during motor movement. J Neurosci. 27:2424–2432.
Miller KJ, Schalk G, Fetz EE, den Nijs M, Ojemann JG, Rao RPN. 2010. Cortical activity during motor execution, motor imagery, and imagery-based online feedback. Proc Natl Acad Sci. 107:4430–4435.
Pantev C, Oostenveld R, Engelien A, Ross B, Roberts LE, Hoke M. 1998. Increased auditory cortical representation in musicians. Nature. 392:811–814.
Pasley BN, David SV, Mesgarani N, Flinker A, Shamma SA, Crone NE, Knight RT, Chang EF. 2012. Reconstructing speech from human auditory cortex. PLoS Biol. 10:e1001251.
Petsche H, von Stein A, Filz O. 1996. EEG aspects of mentally playing an instrument. Brain Res Cogn Brain Res. 3:115–123.
Pitt MA, Crowder RG. 1992. The role of spectral and dynamic cues in imagery for musical timbre. J Exp Psychol Hum Percept Perform. 18:728–738.
Rauschecker JP. 2001. Cortical plasticity and music. Ann N Y Acad Sci. 930:330–336.
Reddy L, Tsuchiya N, Serre T. 2010. Reading the mind's eye: decoding category information during mental imagery. Neuroimage. 50:818–825.
Roth M, Decety J, Raybaudi M, Massarelli R, Delon-Martin C, Segebarth C, Morand S, Gemignani A, Décorps M, Jeannerod M. 1996. Possible involvement of primary motor cortex in mentally simulated movement: a functional magnetic resonance imaging study. Neuroreport. 7:1280–1284.
Saenz M, Langers DRM. 2014. Tonotopic mapping of human auditory cortex. Hear Res. 307:42–52.
Schürmann M, Raij T, Fujiki N, Hari R. 2002. Mind's ear in a musician: where and when in the brain. Neuroimage. 16:434–440.
Sturm I, Blankertz B, Potes C, Schalk G, Curio G. 2014. ECoG high gamma activity reveals distinct cortical representations of lyrics passages, harmonic and timbre-related changes in a rock song. Front Hum Neurosci. 8:798.
Tankus A, Fried I, Shoham S. 2012. Structured neuronal encoding and decoding of human speech features. Nat Commun. 3:1015.
Theunissen FE, Sen K, Doupe AJ. 2000. Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. J Neurosci. 20:2315–2331.
Thirion B, Duschenay E, Michel V, Varoquaux G, Grisel O, VanderPlas J, granfort alexandre, pedregosa fabian, Mueller A, Louppe G. 2011. scikitlearn. J Mach Learn Res. 12:2825–2830.
Tian B. 2004. Processing of frequency-modulated sounds in the lateral auditory belt cortex of the rhesus monkey. J Neurophysiol. 92:2993–3013.
Towle VL, Yoon H-A, Castelle M, Edgar JC, Biassou NM, Frim DM, Spire J-P, Kohrman MH. 2008. ECoG gamma activity during a language task: differentiating expressive and receptive speech areas. Brain. 131:2013–2027.
Wu MC-K, David SV, Gallant JL. 2006. Complete functional characterization of sensory neurons by system identification. Annu Rev Neurosci. 29:477–505.
Yoo SS, Lee CU, Choi BG. 2001. Human brain mapping of auditory imagery: event-related functional MRI study. Neuroreport. 12:3045–3049.
Zatorre RJ, Halpern AR. 1993. Effect of unilateral temporal-lobe excision on perception and imagery of songs. Neuropsychologia. 31:221–232.
Zatorre RJ, Chen JL, Penhune VB. 2007. When the brain plays music: auditory–motor interactions in music perception and production. Nat Rev Neurosci. 8:547–558.
Zatorre RJ, Halpern AR. 2005. Mental concerts: musical imagery and auditory cortex. Neuron. 47:9–12.
Zatorre RJ, Halpern AR, Bouffard M. 2009. Mental reversal of imagined melodies: a role for the posterior parietal cortex. J Cogn Neurosci. 22:775–789.
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.