Decoding Imagined Musical Pitch From Human Scalp Electroencephalograms

Miyoung Chung, Taehyung Kim, Eunju Jeong, Chun Kee Chung, June Sic Kim, Oh-Sang Kwon, and Sung-Phil Kim

***Abstract*— Brain-computer interfaces (BCIs) can restore impaired cognitive functions in people with neurological disorders such as stroke. Musical ability is a cognitive function that is correlated with non-musical cognitive functions, and restoring it can enhance other cognitive functions. Pitch sense is the most relevant function to musical ability according to previous studies of amusia, and thus decoding pitch information is crucial for BCIs to be able to restore musical ability. This study evaluated the feasibility of decoding pitch imagery information directly from human electroencephalography (EEG). Twenty participants performed a random imagery task with seven musical pitches (C4–B4). We used two approaches to explore EEG features of pitch imagery: multiband spectral power at individual channels (IC) and differences between bilaterally symmetric channels (DC). The selected spectral power features revealed remarkable contrasts between left and right hemispheres, low- (<13 Hz) and high-frequency (> 13 Hz) bands, and frontal and parietal areas. We classified two EEG feature sets, IC and DC, into seven pitch classes using five types of classifiers. The best classification performance for seven pitches was obtained using IC and multiclass Support Vector Machine with an average accuracy of 35.68±7.47% (max. 50%) and an information transfer rate (ITR) of 0.37±0.22 bits/sec. When grouping the pitches to vary the number of classes (K = 2–6), the ITR was similar across K and feature sets, suggesting the efficiency of DC. This study demonstrates for the first time the feasibility of decoding imagined musical pitch directly from human EEG.**

***Index Terms*— Decoding, music brain–computer interface, musical pitch, EEG, spectral feature.**

Manuscript received 12 June 2022; revised 30 January 2023 and 11 April 2023; accepted 18 April 2023. Date of publication 25 April 2023; date of current version 3 May 2023. This work was supported in part by the Brain Convergence Research Program of the National Research Foundation (NRF) funded by the Korean Government (MSIT) under Grant NRF-2019M3E5D2A01058328 and Grant 2021M3E5D2A01019542 and in part by the Samsung Research Funding & Incubation Center of Samsung Electronics under Project SRFC-IT1902-08. *(Corresponding authors: Oh-Sang Kwon; Sung-Phil Kim.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board (IRB) of the Ulsan National Institute of Science and Technology under Application No. UNISTIRB-20-22-A.

Miyoung Chung is with the Department of Biomedical Engineering, Ulsan National Institute of Science and Technology, Ulsan 44919, Republic of Korea, and also with the Montreal Neurological Institute, McGill University, Montréal, QC H3A 2B4, Canada (e-mail: miyoung.chung@mail.mcgill.ca).

Taehyung Kim, Oh-Sang Kwon, and Sung-Phil Kim are with the Department of Biomedical Engineering, Ulsan National Institute of Science and Technology, Ulsan 44919, Republic of Korea (e-mail: kth9934@unist.ac.kr; oskwon@unist.ac.kr; spkim@unist.ac.kr).

Eunju Jeong is with the Department of Music Therapy, Graduate School, Ewha Womans University, Seoul 03760, Republic of Korea (e-mail: ejeong@ewha.ac.kr).

Chun Kee Chung is with the Department of Brain and Cognitive Science, the Department of Neurosurgery, the College of Natural Science, and the College of Medicine, Seoul National University, Seoul 08826, South Korea, and also with the Department of Neurosurgery, Seoul National University Hospital, Seoul 03080, South Korea (e-mail: chungc@snu.ac.kr).

June Sic Kim is with the Department of Brain and Cognitive Science, the Department of Basic Sciences, and the College of Natural Science, Seoul National University, Seoul 08826, South Korea (e-mail: jskim@hbf.re.kr).

This article has supplementary downloadable material available at https://doi.org/10.1109/TNSRE.2023.3270175, provided by the authors.

Digital Object Identifier 10.1109/TNSRE.2023.3270175

I. Introduction

THE advance of brain-computer interfaces (BCIs) has facilitated direct interactions between the human brain and the outer world. Among the different types of BCIs, active BCIs have been proven to restore various functions by reinforcing congruent brain activity features during imagery tasks [1]. To develop active BCIs, it is crucial to detect key features of the brain activity elicited from the imagery task in the first place [1]. Electroencephalogram (EEG)-based motor imagery BCIs (MI-BCIs), which harness the features of brain activity generated by imagining movements, are the most extensively used active BCIs. MI-BCIs have been shown to restore impaired motor function in patients suffering from neurological disorders such as post-stroke syndrome, spinal cord injury, and disorders leading to locked-in syndrome by recovering damaged neuronal circuits through neuronal plasticity [2], [3], [4], [5]. Cognitive functions, including language, attention, and memory, are also prominently impaired in patients with post-stroke syndrome, termed post-stroke syndrome cognitive impairment (PSCI) [6]. Currently, BCI-based training methods for PSCI have been recognized for their reduction of cognitive anomalies compared to traditional training methods [2], [7].

A deficit in musical ability is one form of PSCIs, expressed as acquired amusia (AA) in patients who experience brain lesions predominantly in the right superior/middle temporal gyrus (STG/MTG), inferior frontal gyrus (IFG), and hippocampus; importantly, restoring musical ability appears to be crucial for the restoration of other cognitive functions

[8], [9]. Musical ability is well known for its correlation with myriad non-musical cognitive functions such as language, intelligence, memory, and attention [9], [10], [11]. Furthermore, musical ability is regarded as more fundamental than linguistic ability because patients with severe dementia lose their linguistic but not musical ability, suggesting the possibility of new communication channels for patients with defective linguistic functions [12]. Therefore, rehabilitation of musical ability could positively contribute to the restoration of cognitive functions and communicative aids in patients with neurological disorders [9], [10], [11], [12].

One of the primary musical factors is pitch, defined as the auditory sensation ordered from 'low' to 'high,' revealed to affect all-inclusive musical ability by studies of amusia (i.e., tone-deafness). There are two broad types of amusia: acquired amusia (AA) and congenital amusia (CA). AA manifests as brain damage caused by neurological disorders such as stroke, notably in the right hemisphere, affecting ventral and dorsal connectivity [14]. In particular, brain damage affecting ventral connectivity between the right temporal and inferior frontal areas with core lesions in the insula and striatum appears to be crucial for AA to develop [8], [14]. CA is an inherent disorder of comprehensive musical ability and the principal deficit is thought to be pitch processing [15]. A behavioral study revealed that the inability to detect pitch changes was dominant in CA patients among various musical ability measures, showing the authority of pitch processing in comprehensive musical ability [15]. Supporting these behavioral results, neuroimaging studies have reported that lesions in the right frontotemporal cortical networks, eminent for pitch processing in the early developmental stage, have been found in CA [16], [17]. In both behavioral and relevant brain connectivity, the reinforcement of pitch processing has been shown to have a potential role in the recovery of amusia in recent studies; specifically, the right dorsal connectivity has been shown to be key to the recovery of AA, and longitudinal training in discriminating pitch and melody enhances the musical ability of individuals with CA [14], [18], [19]. Taken together, these results suggest that the musical ability of amusics is enhanced by strengthening brain networks related to pitch processing.

Given the accomplishments of recent speech BCI studies, the restoration of defective musical functions by active BCIs for patients with amusia could be achievable by decoding pitch-imagery-inducing brain patterns using motor-related strategies designed for BCI training [1], [2]. Recently, Anumanchipalli et al. built a speech BCI by decoding spoken sentences from sensorimotor cortical activity via high-density electrocorticography (ECoG) signals [20]. Moses et al. enabled patients with anarthria to type a sentence in real time by decoding covertly spoken words from sensorimotor cortical activity acquired from subdural ECoG [21]. These achievements of active speech BCIs, based on a motor-auditory integration strategy, indicate the possibility of successful active music BCIs, given the similarities between music and language. It is known that music and language share neural pathways, and amusia and aphasia are related to dysfunctions of the right ventral stream and neural substrates, including the bilateral precentral gyrus and superior temporal plane related to semantic and melody processing [8], [22]. Furthermore, a behavioral study reported that pitch processing shows the most solid correlation between language and music, which postulates a compelling role of pitch construction in music and language: pitch builds intonation and semantic differences in language, and melody in music [23], [24]. Thus, following the path of speech BCIs, starting with pitch imagery-based BCIs, could lead to the realization of active music BCIs.

Many studies have attempted to decode musical information from brain activity but over a wider scope than single pitches. Schaefer et al. decoded seven segments of classical and contemporary music based on temporal EEG patterns [25]. A study of a tonal hierarchical representation of pitch decoded two classes of tonal relationships–in-key/out-of-key, tonic/dominant, or minor 2nd/augmented information–from magnetoencephalography (MEG) [26]. Another study decoded contextual pitch information by classifying the position where the same pitch (440 Hz) was presented with a lower pitch (110 Hz) or higher pitch (1,760 Hz) using brain connectivity features of EEG [27]. However, no study has reported the decoding of individual musical pitches directly from brain signals. Consequently, the feasibility of decoding imagery of individual pitches from brain activity remains elusive.

Understanding neural representations of pitch is crucial for scrutinizing the brain activity features of pitch imagery. Although brain activity for pitch imagery remains less explored [28], capitalizing on perceptual pitch processing in the brain could be insightful, as the imagery and perception of sound are known to share neural networks incorporating the secondary motor area and dorsal premotor cortex [29]. In musical processing, perception and imagery share frontal and temporal cortical regions in the right hemisphere [30], [31]. Although the neural processing of single musical pitch perception in humans requires further investigation, right lateralization is reported to be an essential attribute in pitch perception. For example, patients with right lateral Heschl's Gyrus (HG) resection exhibit a shortfall in the detection of pitch change direction, which supports the hypothesis that the right HG plays the role of a 'pitch centre' for humans [14], [32]. In addition, the right temporal and frontal cortices are involved in melody perception when an active pitch memory task is performed [33]. In contrast, other studies have revealed the relevance of the left hemisphere to melody contour in synchronization with other brain areas, implying the necessity of broader brain inspection beyond right lateralized neural responses in musical pitch processing [34], [35].

This study investigated the feasibility of decoding individual musical pitches (C4, D4, E4, F4, G4, A4, and B4) directly from human brain activity when attempting to cover each pitch covertly by discovering pitch-related spectrotemporal features from EEG recordings. We embrace a comprehensive method for exploring all possible EEG features from the bilateral hemispheres as well as a subtractive method apprehending spectral power differences between paired channels of EEG across bilateral hemispheres to capitalize on the

property of hemispheric asymmetry for musical pitch processing. Focusing on the spectral feature space with five frequency bands (delta, theta, alpha, beta, and gamma), we devised a heuristic and automatic method to compare all spectral power differences across all pitch pairs to capture the features that most differentiated the seven musical pitches. Here, a variety of classifiers–Naïve Bayes' classifier, multiclass Support Vector Machine (SVM), Linear Discrimination Analysis (LDA), XGBoost, and Long Short-Term Memory (LSTM) models–were implemented to find the optimized model for our purpose. Assuming that musically trained people would generate more distinguishable EEG patterns with better musical imagery capacity than non-trained people [36], we also compared the decoding performance between two participant groups, one with musical training and the other without it. This assumption originated from previous studies on MI-BCIs, whose performance had a positive correlation with the spatial imagery capacity of subjects, hypothesizing the cruciality of individual imagery capacity for achieving better performance in BCIs [37]. We anticipated that unveiling the feasibility of decoding imagined musical pitch from brain activity could help realize a music BCI.

II. Methods

A. Subjects

Twenty-one subjects were recruited. Subjects who never received formal musical training or received less than 3 years of training were allocated to the non-trained (NT) group, and those who received more than 3-years of musical training and met the criteria in the pitch detection ability test (see II.B for details) were allocated to the musically trained (MT) group. Ten subjects were allocated to the MT group (5 females, average age of 24.2±1.33 years) and 10 to the NT (4 females, average age of 25.5±1.84 years), where 1 subject musically trained over 3 years failed the test and was excluded.

All subjects in the MT group were able to play the piano. None of the subjects reported any abnormalities in hearing or brain function. Written consent was obtained from all the subjects before the experiment, and the participants were paid for their participation. This study was approved by the Institutional Review Board (IRB) of the Ulsan National Institute of Science and Technology (UNISTIRB-20-22-A).

B. Stimuli

The auditory stimuli were seven pitch sounds of piano in the 4^{th}-octave musical scale: C4, D4, E4, F4, G4, A4, and B4, to include the international standard note of A4 [38]. The frequencies corresponding to each pitch were 261.63, 293.66, 329.63, 349.23, 392.00, 440.00, and 493.88 Hz, respectively. Each stimulus had a duration of 500 ms. The sound intensity was adjusted to the individual comfort level of each participant, as assessed by their verbal responses. All stimuli were generated using Logic Pro (Apple Inc., Cupertino, CA, USA). A visual stimulus was designed as a piano keyboard image of one scale where the seven-note names in Korean were tagged. During the experiment, a visual stimulus was used to indicate the target pitch of the imagery task (see Section C for more details).

C. Experimental Task

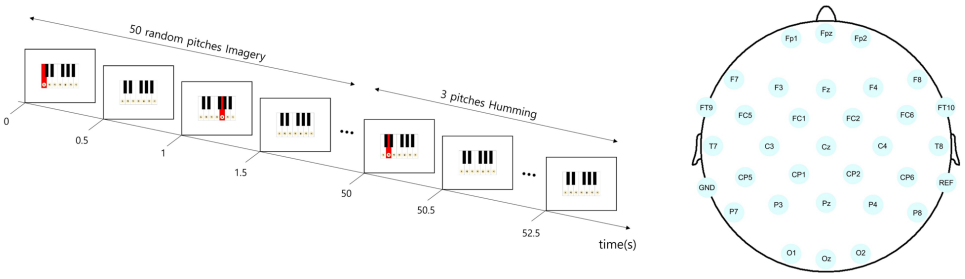
The subjects performed two different tasks during the experiment: a perception task followed by an imagery task. In the perception task, the subjects were asked to count the target pitch among 50 random pitch sounds consisting of seven pitches. First, the subjects were informed of the target pitch, and then 50 pitches were presented in series with a 500 ms inter-stimulus interval (ISI). Each of the seven pitches were randomly presented 50 times, and the subjects counted the number of target pitches and responded to the counted number with the keyboard. This round of perceiving 50 pitches and counting the target pitch was termed a block. The subjects performed a total of 14 blocks, and all seven pitches were set as the target pitch twice. We pseudo-randomized the stimulus presentation and acquired 100 trials for each pitch sound over 14 blocks. During the perception task, an image of the piano keyboard without any cues was displayed on the computer screen to help the subjects concentrate on the musical scale. The performance of the perception task determined a final MT group allocation (see Section II-A); if subjects with >3 years of musical training counted the target pitch correctly in 10 blocks out of 14, they were in the MT group; otherwise, the subject was excluded from the experiment. This exclusion criterion was not applied to subjects without musical training.

In the imagery task, subjects subvocalized the humming of pitch tones on a musical scale (Fig. 1A). There were also 14 blocks in the imagery task, where block construction was the same as the perception task but without sound and target pitch per block. At the beginning of each block, the subjects initially listened to ascending sounds from C4 to B4 for a mental representation of the musical scale. The subjects then performed the task with the guidance of a visual cue on the image of a piano keyboard. The visual cue randomly turned one of the seven keys red for 500 ms, followed by an ISI of 500 ms, and subjects imagined the humming of the cued pitch. The subjects performed 50 trials per block without any auditory stimulus. After each block, the subjects overtly hummed the 3 cued pitches selected at random to monitor how well they were tracking the task. There was an apparent difference in overt humming performance between the groups; the NT group had lower precision but a higher response time than the MT group. We acquired 100 trials for each pitch over the 14 blocks. The subjects were allowed to rest after each block for no longer than 2 min.

The experimental paradigm was implemented using MATLAB Psychtoolbox (Mathworks, Inc., Natick, MA, USA). Visual stimuli were presented on an LCD monitor of 1920×1080 resolution, and auditory stimuli were presented through earphones plugged into both ears. In this study, only data from the imagery task were analyzed according to the study goal.

D. EEG Data Acquisition and Processing

EEG signals were acquired using an EEG amplifier (Acti-Champ, Brain Product GmbH, Gilching, Germany) with 31 active wet electrodes (FP1, FPz, FP2, F7, F3, Fz, F4, F8, FC9, FC5, FC1, FC2, FC6, FC10, T7, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, O1, Oz,



(A) (B)

Fig. 1. (A) A pitch imagery task. The pitch imagery task included 50 trials of pitch imagery. In each trial, a visual cue (red keyboard) appeared randomly on one of the seven notes (C4–B4) for 0.5 s, cueing participants to imagine the corresponding pitch covertly. A 0.5-s inter-stimulus interval followed with no cued keyboard. After 50 trials, participants hummed overtly following visual cues randomly presented for 3 times. (B) EEG channel Montage. Thirty-one EEG electrodes were placed at the locations following the 10/20 international standard.

and O2) at a 500-Hz sampling rate (Fig. 1B). Ground and reference electrodes were placed on the mastoids of the left and right ears, respectively, following the 10-20 system of the American Clinical Neurophysiology Society guideline 2 [39]. The impedance of the electrodes was kept below 10k-Ohm except for a few electrodes (3 at maximum), where any impedance did not exceed 20k-Ohm.

The preprocessing of EEG signals was as follows: 1) the high-pass filter with a 1-Hz cutoff was applied to the EEG signal; 2) the line noise was removed by a notch filter at 60 Hz with a 2-Hz bandwidth; 3) the band-pass filter with a passband from 1 Hz to 50 Hz was applied to the signal; 4) bad channels were detected and removed as follows: signals were low-pass filtered ($<$1 Hz), then the channels were scrutinized piece-wisely, and a channel was judged to be a bad channel if it had a lower cross-correlation than 0.4 across more than 70% of the total channel [40]; 5) The common average reference (CAR) technique removed a potential common noise component from the diverse reference selections [41]; and 6) the artifact subspace reconstruction (ASR) method with a cutoff of 30 eliminated the artifacts, which is reported to preserve brain activity and remove artifacts [42]. Note that this artifact removal aimed to refine the EEG signal rather than dropping out bad trials. Subsequently, a balanced number of trials, 100 trials for each pitch class was guaranteed. All preprocessing was implemented using EEGLAB software [43].

The preprocessed EEG signals were epoched from -200 to 1,000 ms, based on the stimulus onset. In each epoch, event-related spectral perturbation (ERSP) was extracted as follows: First, the epoched signal at each channel was transformed to time-varying spectral power data via short-time Fourier transform (STFT) with a 100-ms sliding window and 90% overlap by the *spectrogram()* function in the MATLAB Signal Processing Toolbox. By setting the number of points for the discrete Fourier transform computation to 512, the frequency resolution was 0.98 Hz, therefore, the frequency bin size was approximately 1 Hz. The number of frequency bins (f) was 50 and the number of time samples (t) was 111 for the transformed STFT, yielded by the passband range at 3) and by the size of the sliding window and overlapping, respectively. The resulting STFT for one trial was $X_{STFT} \in R^{f\times t\times CH} = R^{50\times 111\times 31}$, where CH was the total number of channels. Baseline correction was employed for every frequency bin by dividing each spectral power value of the bin by the baseline mean value averaged from -200 ms to 0 ms. Then, the baseline-corrected values were transformed to the log scale. Five frequency bands were generated by averaging the baseline-corrected spectral power values in the following frequency ranges: delta, 1–4 Hz; theta, 4–8 Hz; alpha, 8–13 Hz; beta, 13–30 Hz; and gamma, 30–50 Hz. The averaged spectral power values in each band were normalized over time using the z-score method. As we intended to extract imagined pitch information guided by the cue, 0–800 ms after stimulus onset, we collected band power values after stimulus onset with 81 time points out of 111 points. Consequently, the matrix for each trial was $X_{input} \in R^{5\times 81\times 31}$ and was submitted to the subsequent feature extraction procedure.

We assigned the class label to each trial according to the pitch information (a total of seven classes). We arranged the trials for each pitch class in chronological order: the first 80% of trials in a training set, and the last 20% in a test set for each class. Note that the decision to drop the cross-validation scheme here was because we opted to arrange the training and test sets in a fashion similar to the practical operation of BCIs, where a training set is collected and used to build decoding models before testing BCIs. As such, there were 560 trials in the training set, $X_{train} \in R^{5\times 81\times 31\times 560}$, and 140 trials in the test set, $X_{test} \in R^{5\times 81\times 31\times 140}$, respectively.

E. Feature Extraction

We explored the features that distinguish the seven pitches from the spectral power distribution over time for every frequency band and channel from the training set. This exploration embodied two different schemes: using the spectral power values of all individual channels (IC) and differences between bilateral channels (DC).

In the IC scheme, we probed the time courses of spectral power averaged over trials for each class at every frequency band and channel (e.g., see Fig. 2A–B). From visual inspection, we observed that the time courses of each class crossed at a certain time point after stimulus onset and then diverged. Such divergence of the time courses appeared to be maximal immediately after the convergence. This pattern was observed in most of the bands and channels (Fig. S1). Based on these

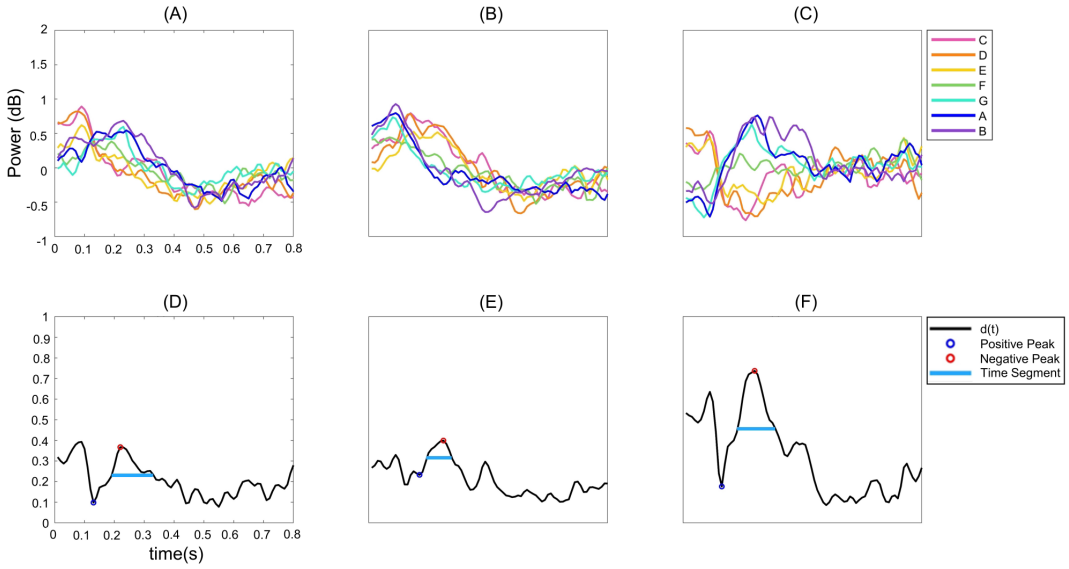


Fig. 2. Feature extraction procedure. The time courses of the mean theta band power from the visual cue onset (0 s) to 0.8 s after the onset for each of the seven pitches (C4–B4) are depicted for F3 (A) and F4 (B) EEG channels, respectively. Difference between the two channels (F3 - F4) are also depicted (C). The divergence of the time courses across pitches (d(t)) for each set of the time courses at F3 (D), F4 (E), and their difference (F) is depicted. The positive (blue circle) and negative (red circle) peaks used for time segment selection, and the resulted time segment (cyan line) are marked. Note that the power values at 0 s of (A)–(C) is varied from 0 dB as they were normalized independently after the baseline correction.

observations, we devised a metric to assess the divergence of the time courses according to the pitch at each time instant, as follows:

$$d\,(t) = \frac{1}{P}\sum_{i}^{P} |x_{i1}\,(t) - x_{i2}\,(t)| \tag{1}$$

where $d(t)$ is the mean of the pairwise absolute differences between all pairs of seven pitches in band power at time t, and x_{i1}(t) and x_{i2}(t) are the averaged band powers at time t of the i-th pair of pitches for $i = 1, \ldots, P$, where P is the number of pitch pairs ($P = 21$). First, we identified the negative and positive peaks of $d(t)$. Among these peaks, we inspected a pair of negative and positive peaks, where the negative peak preceded the positive peak, and selected the pair that showed the largest difference between peaks that could reflect the crossing followed by divergence of the time courses. The selected peaks were used to calculate the time segment, which corresponded to the full width at half maximum (FWHM) of the gap between the negative and positive peaks (blue lines in Fig. 2.D–E). Subsequently, a feature was extracted as an area under the time courses of spectral power within the calculated time segment for every trial.

In the DC scheme, we extracted features from hemispheric differences in band power based on the observation that the averaged time courses of band power for each pitch exhibited opposite patterns between the left and right hemispheric channels (see Fig. 2. A–B). To capitalize on this contrast, we subtracted the band power of the right hemispheric channel from the left counterpart for each channel pair, and 13 bilaterally symmetric channel pairs were arranged as follows: Fp1-Fp2, F7-F8, F3-F4, FC9-FC10, FC5-FC6, FC1-FC2, T7-T8, C3-C4, CP5-CP6, CP1-CP2, P7-P8, P3-P4, and O1-O2. Consequently, the crossing and divergence of the time courses of band power shown in the individual channels became more pronounced, as demonstrated in Fig. 2C. From the time courses of the band power, a set of features was extracted in the same way as in the IC scheme using $d(t)$ (Fig. 2F). The selected positive peaks of $d(t)$ in the DC scheme were larger than those in the IC scheme (Fig. S2).

As there were five bands with 31 channels or 13 channel pairs, the size of the feature set was 155 for the IC and 65 for the DC schemes. The features were further evaluated using the one-way ANOVA test and selected when there was a significant difference among the seven classes ($p < 0.01$). The number of selected features was 123.9±16.6 with the IC scheme and 54.9±5.96 with the DC scheme on average across subjects. Note that all feature inspection and statistical tests were implemented with MATLAB built-in functions and the Statistical and Machine Learning Toolbox.

F. Decoding Model

This study compared several classification algorithms to decode the selected features into seven pitch classes, including Naïve Bayes' classifier, multiclass SVM, LDA, XGBoost, and LSTM (see SX for details in models). The performance of the classifiers was evaluated based on the test accuracy, information transfer rate (ITR), and diagonality of the confusion matrix.

Test accuracy was defined as the ratio of the number of corrected trials to the total number of trials in the test set. ITR was calculated as follows [44]:

$$B = \log_2 N + A\log_2 A + (1-A)\log_2 \frac{1-A}{N-1} \tag{2}$$

where N is the number of classes and A is the test accuracy. B was divided by 0.8 sec, obtaining the ITR unit as bit/s.

The diagonality of the confusion matrix was calculated to examine how close the misclassified pitch was to the true one. As the pitch classes were linear, misclassification into a closer pitch could be regarded as better than a further

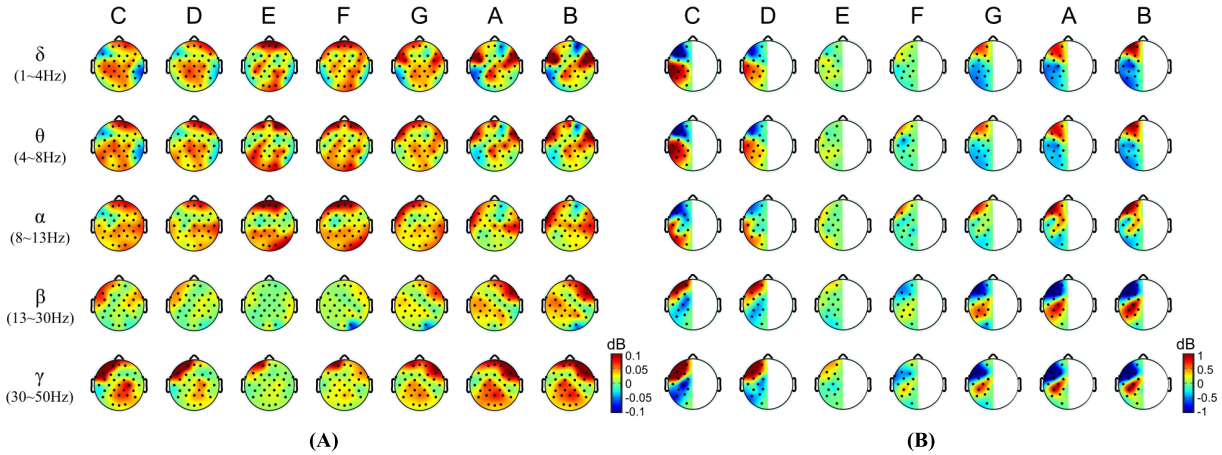


Fig. 3. Topographic distributions of EEG spectral features. (A) Individual channel (IC) features, and (B) Difference between bilateral channel (DC) features. DC features are displayed on the left hemisphere for convenience.

pitch. For example, if the true class was C4, decoding as D4 would be considered less confusing than decoding as A4, although both were treated as misclassification in terms of accuracy. We methodized this examination as the diagonality of a confusion matrix by correlating a confusion matrix with a semi-diagonal matrix, where the semi-diagonal matrix contained two diagonal and one adjacent-to-diagonal entries (Fig. S3). We calculated the 2-D correlation (r) between these two matrices as follows:

$$r = \frac{\sum_m \sum_n (S_{mn} - \bar{S})(C_{mn} - \bar{C})}{\sqrt{(\sum_m \sum_n (S_{mn} - \bar{S})^2)(\sum_m \sum_n (C_{mn} - \bar{C})^2)}} \quad (3)$$

where S is the semi-diagonal matrix, and C is a confusion matrix. S_{mn} (C_{mn}) is an entry of S (C) at the m-th row and n-th column, and $\bar{S}$ ($\bar{C}$) is the mean of all entries in S (C).

Not only did we decode seven individual pitches, but we also decoded a group of pitches to inspect if the decoding performance varied by the number of classes. We grouped pitches into K classes ($1 < K < 7$) under the following conditions: the groups must be chunked with pitches linearly adjacent to each other, and the number of pitches per group must be as balanced as possible. For example, with $K = 3$, pitches can be grouped as CDE/FG/AB, CD/EFG/AB, or CD/EF/GAB, but not as CDG/EA/FB (pitches are not adjacent) or C/DEFG/AB (not the most balanced). We explored all possible cases of the groupings, and those with the highest classification accuracy with all classifiers were determined as the final grouping for each K.

III. Results

In this section, the extracted features for both the IC and DC schemes are first presented. We then report the best decoding results for K classes ($K = 1, 2, \ldots, 7$) using either the features extracted via the IC scheme (IC features) or the DC scheme (DC features). Finally, we compared the decoding outcomes of the MT and NT groups.

A. Feature Distribution

We explored the spatial distributions of the IC and DC features obtained from the training set, as shown in Fig. 3 (for a representative subject [subject 11]). For the IC features, we observed systematic changes in the spatial distribution according to the pitch height (Fig. 3A). In the lower frequency bands (delta, theta, and alpha bands), the higher feature value distribution in the frontal region gradually migrated from the right to left hemispheres as the pitch height increased. In the higher frequency bands (beta and gamma bands), the opposite migration of larger feature values in the frontal region was observed from the left to right hemispheres. Meanwhile, in the temporoparietal region, such migrations showed a reversed propensity: migration of larger features from left to right in the lower bands, and from right to left in the higher bands.

The characteristics of the spatial distributions of features according to the pitch height were evinced more vividly in the DC features (Fig. 3B), which aligns with the results shown in Fig. S2. Note that Fig. 3B depicts the feature differences between hemispheres on the left hemisphere in relation to visualization, and a channel difference was calculated by subtracting the right hemispheric feature values from the left counterparts. We observed conspicuous contrasts in the DC feature distribution pursuant to the pitch height between the frontal and temporoparietal regions, as well as between low- and high-frequency bands, with a more apparent interaction between brain region and frequency. The DC feature values increased as the pitch height increased in the frontal region and low-frequency bands (delta, theta, and alpha) or in the temporoparietal region and high-frequency bands (beta and gamma). In contrast, they decreased as the pitch height increased in the frontal region and high-frequency bands or in the temporoparietal region and low-frequency bands.

The penchant of the spatial distributions of the IC or DC features largely remained after selection with one-way ANOVA across single trials in the training set (Fig. S4), and were more similar between adjacent pitches. Notably, the spatial distributions of the features appear to be clustered into {C, D}, {E, F, G}, and {A, B}. The selected feature distribution for all subjects is depicted in Fig. S5 with the trials averaged by pitch classes.

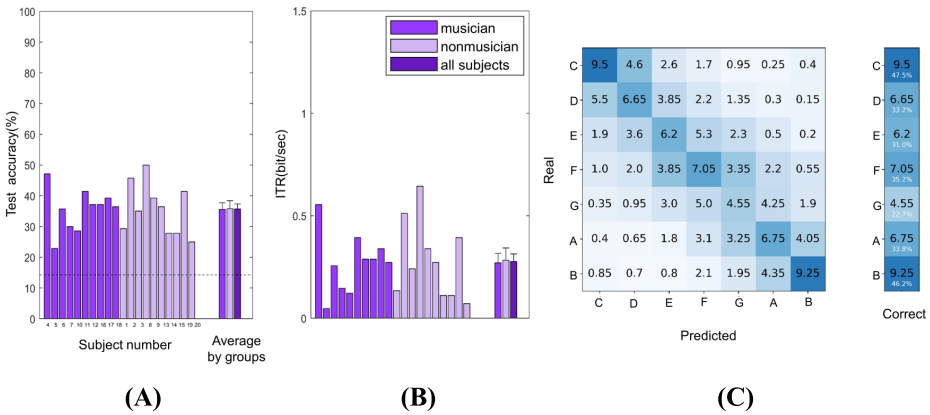


Fig. 4. Decoding performance. The deciding performance of seven pitches from EEG obtained by the best combination of a feature set (IC features) and a classifier (SVM) is illustrated in terms of (A) Accuracy, (B) ITR, and (C) Confusion Matrix for individual subjects, by MT and NT groups, and average of those.

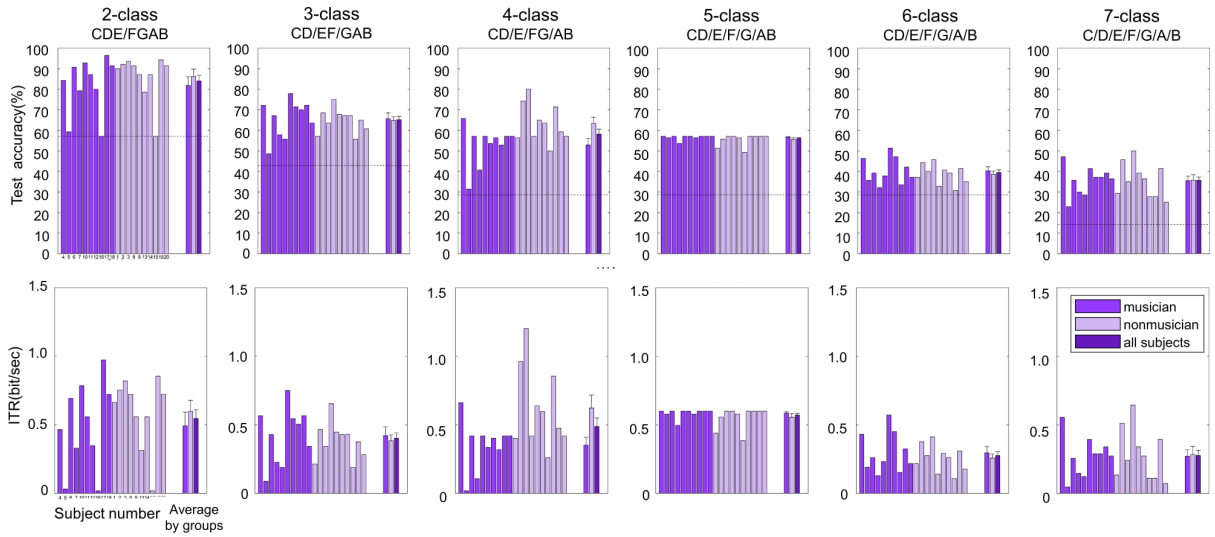


Fig. 5. Decoding performance for a different number of classes of pitch. Decoding performance for all K classes (K = 2, 3, . . . , 7) is illustrated in terms of (top) Accuracy and (bottom) ITR. The chance level is depicted as a black dashed line, defined by the maximum chance of the model where they predict most in the random, maximum number of pitches in the grouping divided by K. The best combination of feature and classifier is DC feature & LSTM for 2, 4, and 5 classes, IC & LSTM for 3 classes, and IC & SVM for 6 and 7 classes.

B. Decoding Individual Pitches

First, seven individual pitches were decoded from the IC or DC features. We classified the features extracted from the test set using five classifiers (Table S1 and below for details). We then compared the classification accuracy and ITR among the classifiers and between the feature schemes (IC vs. DC) using the Scheirer-Ray-Hare (SRH) test, a non-parametric 2-way ANOVA test. There was no main effect of the classifier or interaction effect ($p>0.05$), but there was a significant difference for feature schemes ($p<0.01$). A post-hoc analysis using the Kruskal-Wallis (KW) test revealed that the IC feature yielded better performance in accuracy and ITR than the DC ($p<0.05$). The best decoding performance was achieved by using IC with SVM, resulting in an average accuracy of 35.68±7.47% (max. 50%) and an average ITR of 0.28±0.16 bits/sec (Fig. 4). In addition, we compared the computation time taken to run decoding 7 classes by SVM for each feature set, 4.6±0.7 ms with IC and 3.5±0.3 ms with DC, indicating the time-savings associated with the DC (Table. S2).

We constructed the confusion matrix from the classification outcomes of the five classifiers for all 2 feature schemes (Fig. S6–7). The SRH test showed the main effect of the classifiers ($p<0.05$), followed by the Dunn test, which showed that LSTM yielded lower diagonality than the others ($p < 0.05$; Fig. S8).

C. Decoding Groups of Pitches

Classifications into K pitch classes were evaluated for each K: $1< K <7$. The grouping results for the K pitch classes are shown in Fig. 5 and Table S3.

We calculated the accuracy and ITR for each K using either the IC or DC with each of the five classifiers (Table S1), and compared the feature scheme and classifiers using the SRH test. For $K = 2$ and $K = 5$, the test showed the main effect of the classifier, and the Dunn test revealed that LSTM yielded the highest accuracy ($p<0.05$). For $K = 3$, a main effect of the feature scheme was observed, and a higher accuracy with the IC was revealed ($p<0.05$). For $K = 4$, the main effects for both the classifier and feature schemes were shown,

and post-hoc analyses revealed higher accuracy using LSTM with the IC feature ($p<0.05$). For $K = 6$, no significant differences were found between the classifiers and feature schemes ($p>0.05$). No interaction effect was observed for any K classes. We optimized the decoding models for each K as a combination of feature schemes and classifiers based on the statistical test results. If there was no main effect of the feature, we selected the DC because it required a smaller feature size than the IC to achieve a similar level of performance. If there was no main effect of the classifier, the best classifier was selected, which could be most effectively implemented both in accuracy and computation time. The LSTM was not selected unless it showed significantly higher accuracy than others, owing to its much longer computation time. The decoding accuracies from the best combinations were 84.07±12.34% (max. 96.43%) for 2 classes, 65.21±7.42% (max. 77.86%) for 3 classes, 58.18±10.75% (max. 80%) for 4 classes, 56.11±2.17% (max. 57.14%) for 5 classes, and 39.5±5.53% (max. 51.43%) for 6 classes (Fig. 5, top). All of these accuracy values were significantly higher than the corresponding chance levels, calculated as a maximum number of pitches in the grouping divided by K (t-test, $p<0.05$).

The decoding outcomes of selected decoding models for each K were evaluated via ITR, a consistent measure of decoding performance by considering the different number of classes (Fig. 5, bottom). For all feature sets, there was no interaction between the classification model and the number of classes (K) (SRH test, $p>0.05$), but the number of classes showed the main effect. The effect of K was tested for each feature type using the Dunn test; the ITR for 5 classes showed a relatively higher value. Remarkably, the ITR for 5 classes was significantly higher than that for 2, 6, and 7 classes with IC ($p<0.05$, Fig. S9A) and for 3, 4, 6, and 7 classes with DC ($p<0.05$, Fig. S9B).

The feature distributions in the training set, according to the pitch groups for each K, were visualized using t-SNE (Fig. S10). The DC features were distributed more linearly with pitch height, in contrast to the distribution of IC features that were more spread, suggesting a better decoding result of the IC feature for the 7-class classification.

D. Comparison of Musically Trained and Non-Trained Groups

We evaluated whether the decoding performance was different between the MT and NT groups using the decoding models selected for each K, as described above (see Section III-C). Dunn's test revealed no significant difference in both accuracy and ITR between the groups for all K classes ($p>0.05$). No significant differences in diagonality were found between the groups for all K classes ($p>0.05$).

IV. Discussion

In this study, we decoded the pitch imagery information from EEG by extracting the most discriminable features from the temporal patterns of spectral power in every channel and frequency band. We designed two schemes for feature extraction: individual channels (IC) and differences between channels (DC). Differences were observed between bilateral channels located symmetrically in each hemisphere, according to our observation of a symmetrically reversed feature distribution across hemispheres (Fig. 3). We used each of the IC or DC feature sets to decode pitch with five classifiers and then selected the combinations for each classification of K classes ($2 \leq K \leq 7$) best in both statistical and computation time (Table S1). The classification accuracy was significantly higher than the chance level for every K, although there was room for substantial improvement prior to its application to real-time BCIs. Between the feature sets, using IC was better in terms of the classification accuracy of multiple pitch groups (i.e., large K), whereas DC was better in representing higher or lower pitch (e.g., C4 or B4). However, the ITR of both features showed no difference for multiple pitch groups, suggesting that using the DC was effective in representing pitch height information, and even efficient considering the feature dimension where the DC is half of the IC.

The feature distribution revealed noticeable countering patterns between 1) left and right hemispheres, 2) anterior vs. posterior areas, and 3) low- vs. high-frequency bands. Possible hypotheses for these contrasts are proposed with some neurological basis. First, the bilateral contrast (1) may be related to the temporal sensitivity difference between the hemispheres, leading to different spectral resolutions for each hemisphere [36]; that is, the relative pitch height likely formed a bilateral alignment of features across the hemispheres. A potential neural substrate for the observed anterior-posterior contrast (2) can be conjectured by frontoparietal networks related to pitch discrimination, although the source of distribution patterns remains unexplained [35], [45]. The frequency-dependent distribution (3) can be assumed from the characteristics of different EEG oscillations tracking the acoustic properties of auditory stimuli, as delta to alpha oscillations reflect attentional and acoustic input variations in speech prosodic structures [46].

The commonality of the selected features among subjects was further investigated by counting the number of subjects whose features were selected in each channel and frequency band (Fig. S11). Note that only IC was employed to scrutinize the whole brain distribution of the selected feature. The distribution of commonly selected features was observed over the bilateral frontotemporal and parietal areas, particularly in the lower-frequency bands. These areas correspond to frontotemporal and frontoparietal networks admissible for the recovery of pitch sensation in patients with AA and CA [8], [14]. The features in the low-frequency bands were frequently selected from lateral areas rather than medial areas, implying pitch information processing pathways over lateral areas, as reported in previous studies [47]. The features in the high-frequency bands showed a more complicated distribution, suggesting that the inspection strategy to capture hemispheric differences is more meticulous than the current method of subtracting the counterpart electrodes from left to right.

We examined five classification models, including relatively more advanced models such as LSTM and XGBoost, expecting superior performance. Unfortunately, these models did not outperform the simpler models in this study. This may be related to the insufficient number of training samples

for the advanced models [48]. Thus, enlarging the training data size would enhance the performance of advanced models, thereby increasing the possibility of a more practical pitch-imagery BCI.

The analysis of the MT and NT groups was based on the hypothesis that individual music imagery capacity would improve BCI performance [37]. However, no differences in decoding performance and elicitation of pitch-related EEG features were observed between the MT and NT groups. One possible reason might be the insufficient musical context in the stimulus presentation. Another reason might be the limited number of participants for each group, due to difficulties in recruiting not only ordinary participants but also musicians during the coronavirus disease 2019 pandemic. However, this invariant result by the groups may advocate the utilization of imagined pitch decoding regardless of musical training.

Potential interference of horizontal visual stimuli in decoding could be a concern. We examined this by repeating the classification process without the delta band feature containing activity in the 2-Hz band corresponding to the ISI (0.5 s) and found no difference in decoding performance for all K classes (Kruskal-Wallis test, $p > 0.05$). Moreover, a previous study reported that decoding eye movements with a visual angle of 5° from EEG gained accuracy above a chance level only when electrooculography (EOG) signals were used along with EEG [49]. Since we removed artifacts related to eye movements from EEG, it is unlikely that eye movements with even a smaller visual angle of 2.7° between the piano keyboards might influence the decoding results of this study. In addition to eye movements, spatial representation from the posterior parietal cortex (PPC) was considered [50], but the activity from the parietal lobe must be conserved with the cruciality of its network with the frontal lobe for pitch processing [8], [14]. Nevertheless, the extracted feature is related to pitch processing as selected areas were mostly frontotemporal areas regardless of the frequency bands, and decoding performance was preserved even after the eye movements were ruled out. Nonetheless, the horizontal design of visual stimuli should be carefully re-examined in future studies to ensure that decoding is entirely based on pitch imagery brain activity.

Verifying the feasibility of decoding seven pitches on a musical scale from human EEG, the realization of pitch imagery-based BCI appears plausible, nonetheless, further endeavors to guarantee practical BCI realization are needed. To achieve this, reinforcement of EEG features by neurofeedback training would be effective, improving the corresponding EEG features in MI-BCI [51]. Pitch imagery training with proper design of pitch-relevant visual and/or auditory online feedback can be helpful, as shown in recent reports on the advantages of realistic android-based feedback [52], online visual feedback of EEG features [53], and multisensory feedback [54] in MI-BCIs. In addition, decoding performance can enhance by employing advanced deep learning algorithms, as demonstrated by other EEG-BCI studies [55].

Methods of decoding an imagined single pitch from brain signals proposed here and the associated neurofeedback training approach will also contribute to musical imagery and musical learning research as well as building pitch imagery-based BCIs. Musical imagery is well known for its relevance to musical learning based on auditory-motor interaction [47]. As such, if one can decode imagined musical activity from brain and immediately feed decoding output back to learners by auditory signals, it can enhance musical learning by reinforcing the auditory-motor circuit. As the musical pitch is one of the most fundamental musical elements chiefly related to musical ability, a system that decodes single musical pitches would be advantageous in providing neurofeedback for musical learning.

V. Conclusion

This study revealed the feasibility of decoding imagined pitch on a musical scale using human EEG. We found spectrotemporal features that differentiated the multiclass pitches, represented the linearity of pitch height, and ruminated hemispheric differences. We achieved the performance of decoding pitch imagery information from noninvasive brain signals, which could initiate the development of future pitch-imagery-based BCIs for anyone who can represent pitch covertly heedless of the keen pitch sense.

References

[1] S. Steinert, C. Bublitz, R. Jox, and O. Friedrich, "Doing things with thoughts: Brain–computer interfaces and disembodied agency," *Philosophy Technol.*, vol. 32, no. 3, pp. 457–482, Sep. 2019.

[2] R. Mane, T. Chouhan, and C. Guan, "BCI for stroke rehabilitation: Motor and beyond," *J. Neural Eng.*, vol. 17, no. 4, Aug. 2020, Art. no. 041001.

[3] D. J. Leamy et al., "An exploration of EEG features during recovery following stroke—Implications for BCI-mediated neurorehabilitation therapy," *J. NeuroEng. Rehabil.*, vol. 11, no. 1, p. 9, 2014.

[4] S. Saeedi, R. Chavarriaga, and J. del R. Millán, "Long-term stable control of motor-imagery BCI by a locked-in user through adaptive assistance," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 4, pp. 380–391, Apr. 2017.

[5] C. Guger et al., "Complete locked-in and locked-in patients: Command following assessment and communication with vibro-tactile P300 and motor imagery brain–computer interface tools," *Frontiers Neurosci.*, vol. 11, p. 251, May 2017.

[6] J. Hochstenbach, T. Mulder, J. van Limbeek, R. Donders, and H. Schoonderwaldt, "Cognitive decline following stroke: A comprehensive study of cognitive decline following stroke," *J. Clin. Experim. Neuropsychol.*, vol. 20, no. 4, pp. 503–517, Aug. 1998.

[7] S. E. Kober et al., "Specific effects of EEG based neurofeedback training on memory functions in post-stroke victims," *J. NeuroEng. Rehabil.*, vol. 12, no. 1, p. 107, Dec. 2015.

[8] A. J. Sihvonen, T. Särkämö, A. Rodríguez-Fornells, P. Ripollés, T. F. Münte, and S. Soinila, "Neural architectures of music—Insights from acquired amusia," *Neurosci. Biobehav. Rev.*, vol. 107, pp. 104–114, Dec. 2019.

[9] S. Benz, R. Sellaro, B. Hommel, and L. S. Colzato, "Music makes the world go round: The impact of musical training on non-musical cognitive functions—A review," *Frontiers Psychol.*, vol. 6, p. 2023, Jan. 2016.

[10] A. S. Chan, Y.-C. Ho, and M.-C. Cheung, "Music training improves verbal memory," *Nature*, vol. 396, no. 6707, p. 128, Nov. 1998.

[11] J. C. McLachlan, "Music and spatial task performance," *Nature*, vol. 366, no. 6455, p. 520, Dec. 1993.

[12] A. Baird and S. Samson, "Music and dementia," *Prog. Brain Res.*, vol. 217, pp. 207–235, Jan. 2015.

[13] A. Rakowski, "The domain of pitch in music," *Arch. Acoust.*, vol. 34, no. 4, pp. 429–443, 2009.

[14] L. Stewart, K. von Kriegstein, J. D. Warren, and T. D. Griffiths, "Music and the brain: Disorders of musical listening," *Brain*, vol. 129, no. 10, pp. 2533–2553, Oct. 2006.

[15] I. Peretz, "Congenital amusia: A disorder of fine-grained pitch discrimination," *Neuron*, vol. 33, no. 2, pp. 185–191, 2002.

[16] K. L. Hyde, J. P. Lerch, R. J. Zatorre, T. D. Griffiths, A. C. Evans, and I. Peretz, "Cortical thickness in congenital amusia: When less is better than more," *J. Neurosci.*, vol. 27, no. 47, pp. 13028–13032, Nov. 2007.

[17] P. Moreau, P. Jolicœur, and I. Peretz, "Pitch discrimination without awareness in congenital amusia: Evidence from event-related potentials," *Brain Cognition*, vol. 81, no. 3, pp. 337–344, Apr. 2013.

[18] C. Micheyl, K. Delhommeau, X. Perrot, and A. J. Oxenham, "Influence of musical and psychoacoustical training on pitch discrimination," *Hearing Res.*, vol. 219, nos. 1–2, pp. 36–47, Sep. 2006.

[19] K. L. Whiteford and A. J. Oxenham, "Learning for pitch and melody discrimination in congenital amusia," *Cortex*, vol. 103, pp. 164–178, Jun. 2018.

[20] G. K. Anumanchipalli, J. Chartier, and E. F. Chang, "Speech synthesis from neural decoding of spoken sentences," *Nature*, vol. 568, no. 7753, pp. 493–498, Apr. 2019.

[21] D. A. Moses et al., "Neuroprosthesis for decoding speech in a paralyzed person with anarthria," *New England J. Med.*, vol. 385, no. 3, pp. 217–227, 2021.

[22] M. Yu, M. Xu, X. Li, Z. Chen, Y. Song, and J. Liu, "The shared neural basis of music and language," *Neuroscience*, vol. 357, pp. 208–219, Aug. 2017.

[23] T. K. Perrachione, E. G. Fedorenko, L. Vinke, E. Gibson, and L. C. Dilley, "Evidence for shared cognitive processing of pitch in music and language," *PLoS ONE*, vol. 8, no. 8, Aug. 2013, Art. no. e73372.

[24] Y. Li, C. Tang, J. Lu, J. Wu, and E. F. Chang, "Human cortical encoding of pitch in tonal and non-tonal languages," *Nature Commun.*, vol. 12, no. 1, p. 1161, Feb. 2021.

[25] R. S. Schaefer, J. Farquhar, Y. Blokland, M. Sadakata, and P. Desain, "Name that tune: Decoding music from the listening brain," *NeuroImage*, vol. 56, no. 2, pp. 843–849, May 2011.

[26] N. Sankaran, W. F. Thompson, S. Carlile, and T. A. Carlson, "Decoding the dynamic representation of musical pitch from human brain activity," *Sci. Rep.*, vol. 8, no. 1, p. 839, Jan. 2018.

[27] S. Sakamoto, A. Kobayashi, K. Matsushita, R. Shimizu, and A. Aoyama, "Decoding relative pitch imagery using functional connectivity: An electroencephalographic study," in *Proc. IEEE 1st Global Conf. Life Sci. Technol. (LifeTech)*, Mar. 2019, pp. 48–49.

[28] T. L. Hubbard, "Some methodological and conceptual considerations in studies of auditory imagery," *Auditory Perception Cognition*, vol. 1, nos. 1–2, pp. 6–41, Apr. 2018.

[29] S. C. Herholz, A. R. Halpern, and R. J. Zatorre, "Neuronal correlates of perception, imagery, and memory for familiar tunes," *J. Cognit. Neurosci.*, vol. 24, no. 6, pp. 1382–1397, Jun. 2012.

[30] Y. Ding et al., "Neural correlates of music listening and recall in the human brain," *J. Neurosci.*, vol. 39, no. 41, pp. 8112–8123, Oct. 2019.

[31] R. S. Schaefer, P. Desain, and J. Farquhar, "Shared processing of perception and imagery of music in decomposed EEG," *NeuroImage*, vol. 70, pp. 317–326, Apr. 2013.

[32] R. J. Zatorre, "Pitch perception of complex tones and human temporal-lobe function," *J. Acoust. Soc. Amer.*, vol. 84, no. 2, pp. 566–572, Aug. 1988.

[33] R. Zatorre, A. Evans, and E. Meyer, "Neural mechanisms underlying melodic perception and memory for pitch," *J. Neurosci.*, vol. 14, no. 4, pp. 1908–1919, Apr. 1994.

[34] A. D. Patel and E. Balaban, "Temporal patterns of human cortical activity reflect tone sequence structure," *Nature*, vol. 404, no. 6773, pp. 80–84, Mar. 2000.

[35] D. A. Hall and C. J. Plack, "Pitch processing sites in the human auditory brain," *Cerebral Cortex*, vol. 19, no. 3, pp. 576–585, Mar. 2009.

[36] L. Bishop, F. Bailes, and R. T. Dean, "Performing musical dynamics: How crucial are musical imagery and auditory feedback for expert and novice musicians?" *Music Percept.*, vol. 32, no. 1, pp. 51–66, 2014.

[37] C. Jeunet, B. Nkaoua, S. Subramanian, M. Hachet, and F. Lotte, "Predicting mental imagery-based BCI performance from personality, cognitive profile and neurophysiological patterns," *PLoS One*, vol. 10, no. 12, pp. 1–21, 2015.

[38] J. Kennedy, M. Kennedy, and T. Rutherford-Johnson, *The Oxford Dictionary of Music*. Oxford. U.K.: Oxford Univ. Press, 2012.

[39] J. N. Acharya, A. Hani, J. Cheek, P. Thirumala, and T. N. Tsuchida, "American clinical neurophysiology society guideline 2: Guidelines for standard electrode position nomenclature," *J. Clin. Neurophysiol.*, vol. 33, no. 4, pp. 308–311, 2016.

[40] N. Bigdely-Shamlo, T. Mullen, C. Kothe, K.-M. Su, and K. A. Robbins, "The PREP pipeline: Standardized preprocessing for large-scale EEG analysis," *Frontiers Neuroinform.*, vol. 9, p. 16, Jun. 2015.

[41] D. Yao, L. Wang, R. Oostenveld, K. D. Nielsen, L. Arendt-Nielsen, and A. C. N. Chen, "A comparative study of different references for EEG spectral mapping: The issue of the neutral reference and the use of the infinity reference," *Physiolog. Meas.*, vol. 26, no. 3, pp. 173–184, Jun. 2005.

[42] C. Chang, S. Hsu, L. Pion-Tonachini, and T. Jung, "Evaluation of artifact subspace reconstruction for automatic artifact components removal in multi-channel EEG recordings," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 4, pp. 1114–1121, Apr. 2020.

[43] A. Delorme and S. Makeig, "EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J. Neurosci. Methods*, vol. 134, no. 1, pp. 9–21, Mar. 2004.

[44] B. Obermaier, C. Neuper, C. Guger, and G. Pfurtscheller, "Information transfer rate in a five-classes brain–computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 9, no. 3, pp. 283–288, Sep. 2001.

[45] A. J. Sihvonen et al., "Functional neural changes associated with acquired amusia across different stages of recovery after stroke," *Sci. Rep.*, vol. 7, no. 1, p. 11390, Sep. 2017.

[46] F. Bröhl and C. Kayser, "Delta/theta band EEG differentially tracks low and high frequency speech-derived envelopes," *NeuroImage*, vol. 233, Jun. 2021, Art. no. 117958.

[47] R. J. Zatorre, J. L. Chen, and V. B. Penhune, "When the brain plays music: Auditory–motor interactions in music perception and production," *Nature Rev. Neurosci.*, vol. 8, no. 7, pp. 547–558, Jul. 2007.

[48] M. Z. Alom et al., "A state-of-the-art survey on deep learning theory and architectures," *Electronics*, vol. 8, no. 3, pp. 1–67, 2019.

[49] A. N. Belkacem, H. Hirose, N. Yoshimura, D. Shin, and Y. Koike, "Classification of four eye directions from EEG signals for eye-movement-based communication systems," *J. Med. Biol. Eng.*, vol. 34, no. 6, pp. 581–588, 2014.

[50] N. A. Herweg and M. J. Kahana, "Spatial representations in the human brain," *Frontiers Hum. Neurosci.*, vol. 12, p. 297, Jul. 2018.

[51] B. Zoefel, R. J. Huster, and C. S. Herrmann, "Neurofeedback training of the upper alpha frequency band in EEG improves cognitive performance," *NeuroImage*, vol. 54, no. 2, pp. 1427–1431, Jan. 2011.

[52] C. I. Penaloza, M. Alimardani, and S. Nishio, "Android feedback-based training modulates sensorimotor rhythms during motor imagery," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 3, pp. 666–674, Mar. 2018.

[53] X. Duan, S. Xie, X. Xie, K. Obermayer, Y. Cui, and Z. Wang, "An online data visualization feedback protocol for motor imagery-based BCI training," *Frontiers Hum. Neurosci.*, vol. 15, Jun. 2021, Art. no. 625983.

[54] Z. Wang et al., "A BCI based visual-haptic neurofeedback training improves cortical activations and classification performance during motor imagery," *J. Neural Eng.*, vol. 16, no. 6, Oct. 2019, Art. no. 066012.

[55] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: A review," *J. Neural Eng.*, vol. 16, no. 3, Jun. 2019, Art. no. 031001.