

Frontotemporal oxyhemoglobin dynamics predict performance accuracy of dance simulation gameplay: Temporal characteristics of top-down and bottom-up cortical activities

Yumie Ono ^a, Yasunori Nomoto ^a, Shohei Tanaka ^a, Keisuke Sato ^a, Sotaro Shimada ^a, Atsumichi Tachibana ^b, Shaw Bronner ^c, J. Adam Noah ^{d,*}

^a Department of Electronics and Bioinformatics, Meiji University, Kanagawa, Japan

^b Faculty of Care and Rehabilitation (Physiology), Seijo University, Tohoku, Japan

^c ADAM Center, Bouve College of Health Sciences, Northeastern University, Boston, MA, USA

^d Brain Function Laboratory, Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA

ARTICLE INFO

Article history:

Accepted 13 May 2013

Available online 24 May 2013

Keywords:

fNIRS

Dance video game

Multisensory integration

Middle temporal gyrus

Frontopolar prefrontal cortex

ABSTRACT

We utilized the high temporal resolution of functional near-infrared spectroscopy to explore how sensory input (visual and rhythmic auditory cues) are processed in the cortical areas of multimodal integration to achieve coordinated motor output during unrestricted dance simulation gameplay. Using an open source clone of the dance simulation video game, Dance Dance Revolution, two cortical regions of interest were selected for study, the middle temporal gyrus (MTG) and the frontopolar cortex (FPC). We hypothesized that activity in the FPC would indicate top-down regulatory mechanisms of motor behavior; while that in the MTG would be sustained due to bottom-up integration of visual and auditory cues throughout the task. We also hypothesized that a correlation would exist between behavioral performance and the temporal patterns of the hemodynamic responses in these regions of interest. Results indicated that greater temporal accuracy of dance steps positively correlated with persistent activation of the MTG and with cumulative suppression of the FPC. When auditory cues were eliminated from the simulation, modifications in cortical responses were found depending on the gameplay performance. In the MTG, high-performance players showed an increase but low-performance players displayed a decrease in cumulative amount of the oxygenated hemoglobin response in the no music condition compared to that in the music condition. In the FPC, high-performance players showed relatively small variance in the activity regardless of the presence of auditory cues, while low-performance players showed larger differences in the activity between the no music and music conditions. These results suggest that the MTG plays an important role in the successful integration of visual and rhythmic cues and the FPC may work as top-down control to compensate for insufficient integrative ability of visual and rhythmic cues in the MTG. The relative relationships between these cortical areas indicated high- to low-performance levels when performing cued motor tasks. We propose that changes in these relationships can be monitored to gauge performance increases in motor learning and rehabilitation programs.

© 2013 Elsevier Inc. All rights reserved.

Introduction

Functional near-infrared spectroscopy (fNIRS) is a relatively new tool that allows for recording concurrent behavioral and cortical activities. NIRS can be employed as a noninvasive low-cost optical technique for monitoring tissue oxygen saturation, changes in hemoglobin volume and, indirectly, brain/muscle blood flow and muscle oxygen consumption (Ferrari and Quaresima, 2012; Ferrari et al., 2004). The general compatibility of fNIRS with current fMRI data collection and analysis techniques including the use of general linear model and event related

design further increases the appeal of fNIRS as an important tool for functional neuroimaging (Ferrari et al., 2004). fNIRS is also more compatible than fMRI when studying populations of infants, elderly, and patients with psychoneurological problems (Hoshi, 2003). There are several benefits of employing fNIRS over other more traditional brain recording techniques such as fMRI: first, fNIRS allows subjects to behave in a more natural way while undergoing a scan. Next, fNIRS can employ multiple channel recording of the cortex, which can be observed and manipulated through the behavior of subjects in real-time. Finally, the temporal resolution of fNIRS is significantly higher than that of fMRI (Cui et al., 2011).

Video games coupled with fNIRS provide researchers the ability to understand how we perceive, integrate, and effectively interact with our real-world environment. Here, we studied a complex sequential

* Corresponding author at: Yale University Department of Psychiatry, 300 George Street Suite 902, New Haven, CT 06511, USA.

E-mail address: adam@adamnoah.com (J.A. Noah).

task with auditory and visual elements in a rhythmic dance video game that required players to select the correct motor act at the correct time while continuously processing visual and auditory cues for current and future decisions and acts. In a previous set of experiments using this dance video game paradigm, we investigated the behaviors of simple stepping and complex variations of gameplay with multimodal audio-visual input (Tachibana et al., 2011). We found that task complexity played a role in shaping fNIRS cortical signal amplitude. Additionally, we found different temporal response patterns in areas of sensory integration: the superior parietal lobe and the temporal association area. The capability of fNIRS to determine these temporal response patterns and relationships between cortical areas suggests that fNIRS imaging of actual, non-reduced or simplified video gameplay allows for a greater understanding of how the brain dynamically regulates the attentional networks involved in the sensory-motor processing.

There is some knowledge of the neural systems involved in integrated sensory-motor behaviors as imaged in finger tapping and piano playing studies using fMRI (Bangert et al., 2006; Baumann et al., 2007; Beauchamp et al., 2004; Hasegawa et al., 2004; Jäncke et al., 2000; Zatorre et al., 2007). Functional cortical activities recorded in these studies suggest that the temporal association area, including the superior temporal sulcus and medial temporal gyrus (MTG), serves as an area of integration of multimodal auditory and visual cues using bottom-up mechanisms in which the system responds to integrated information from the environment (Beauchamp et al., 2004; Fuhrmann Alpert et al., 2008; Tankus and Fried, 2012; Visser et al., 2012). Additionally, the frontopolar cortex (FPC) may serve to coordinate multiple cognitive processes from association cortical areas and forward plan to produce effective motor sequences in top-down fashion (Ramnani and Owen, 2004; Sakai et al., 2002). We therefore observed the interaction between functional activities in the FPC, as part of the top-down executive system, and the MTG, as part of the bottom-up sensory processing system, during a complex dance simulation gameplay. These regions are particularly well suited for fNIRS as frontal and temporal areas have the shortest scalp-brain distance (Cui et al., 2011).

We hypothesized that activity in the FPC would indicate regulatory mechanisms of top-down motor control; while that in the MTG would be sustained throughout the task to process continuous sensory input in a bottom-up manner. Furthermore, we hypothesized that there would be temporal and pattern differences in the FPC and MTG depending on player proficiency and details of sensory input.

Materials and methods

Subjects

Twenty-six healthy subjects (five females, 23 right-handed) participated in this study (mean age \pm standard error = 26.1 ± 1.7 years). The study was approved by the Ethics Committee of the Meiji University School of Science and Technology and all subjects gave written informed consent for participation. Each subject was without neurological or psychiatric illness and had normal or corrected-to-normal vision. Subjects had various levels of experience playing this dance video game; 10 were frequent players (more than 5 h per week and up to several years of playing), four had previous experience playing the game but abstained for more than two years, and the remaining 12 subjects were naïve to the game. Regardless of the wide range of reported experience, the performances of subjects were determined only by their timing accuracy of dance steps through the experimental sessions (see ‘Data analysis’ for details).

fNIRS measurement

We used a 22-channel fNIRS topography system OMM-3000 (Shimadzu Co., Kyoto, Japan) arranged into a 3×5 optical probe array. The array was mounted on an elastic optode cap and positioned

over the left prefrontal to the temporal lobes (Figs. 1a–b). The lowest and the most anterior optodes were arranged at Fpz of the international 10–10 system (Chatrian et al., 1985), and the lowest optode row was aligned with the line connecting Fpz-T7. Inter-optode distance was 3 cm for each source detector pair. Data were sampled at 7.9 Hz. The optical probe arrays and optodes were tightly fixed to the cap with a chin-strap to minimize displacements between the head surface and optodes during gameplay. This allowed us to measure reliable fNIRS responses as demonstrated previously (Tachibana et al., 2011). Using a 3D digitizer (PATRIOT, Polhemus, Colchester, VT), we obtained coordinates of all probe positions and the anatomical landmark positions (nasion, inion, auricles and Cz) of each subject immediately before data collection.

Subjects played a dance simulation video game, similar to the commercial game Dance Dance Revolution™ (Konami Corp., Tokyo, Japan) in a block-design fashion. Subjects stood on a dance pad equipped with four buttons consisting of up, down, right, and left arrows and played the song ‘Butterfly’ (recorded by SMILE.dk). A forty-seven inch television was positioned 1.2 m in front of the subject, providing auditory and/or visual game cues. A series of arrow-shaped visual cues in the up, down, right, or left directions moved from the bottom of the screen to the top. Subjects responded by pressing the correct button at the correct time with their foot when an arrow reached a response area located at the top of the screen (Fig. 1c). We made several modifications to the game using the open source software clone of DDR, StepMania. First,



Fig. 1. Instruments of the experiment. a. The subject stood on a dance pad with optical probes of fNIRS on her head. Optodes were arranged in the 3×5 array (indicated by a dashed-line area), which was mounted on an elastic optode cap. b. Game screen provided visual cues (arrow signs), which scrolled up from the bottom of the screen to the top. The subject responded by pressing the same arrow button on the dance pad at the correct time with his or her foot when it reached a response area (indicated by a dashed-line box). In the “with music (WM)” condition, the timing of steps was to the rhythm of the background music of the game.

we changed the background color of the game screen depending on the period of the block design to inform the player whether to play or rest (green for the task and red for the rest period). Second, the directions of arrows and these orders were randomized and subjects were required to step with the upbeat to eliminate the effect of familiarity with the commercial version of this game song. Third, we did not use the simultaneous two-button jump feature to avoid motion artifact from head movement during jumping.

Subjects played the game two times each in two alternating conditions of (i) with music (WM) and (ii) without music (no music: NM). A single trial consisted of five 30 s task periods and five 30 s rest periods in alternating blocks. Subjects were instructed to play the game only during the task period and to minimize additional steps which were not instructed on the game screen or any extraneous head and upper body movement as much as possible throughout the whole trial. Basic mechanics of the game were explained to subjects at the beginning and they practiced each condition before acquisition. The total time for these four runs including practice was approximately 40 min. Dance pad sensors detected position and timing information to provide participants with real-time visual feedback, identical to actual game-play. To measure the behavioral performance of the subjects, we counted the number of temporally accurate steps among the 240 steps presented in each trial, at which the correct button was pressed within ± 22.5 ms of the exact timing.

Assuming that the FPC is responsible for maintaining an intention for future action to be executed at the right timing (Burgess et al., 2001), we hypothesized that the FPC may demonstrate higher activation with the condition in which the player had to pay more attention to achieve temporally accurate steps. Therefore, following completion of all gameplay conditions, a post-experimental attentional effort questionnaire was administered. We asked subjects which condition required more attentional effort to produce accurate steps: (i) WM condition, (ii) NM condition, or (iii) comparable effort.

Data analysis

A modified Beer–Lambert approach (Cope et al., 1988) was used to calculate signals reflecting the oxygenated hemoglobin (oxyHb), deoxygenated hemoglobin (deoxyHb), and total hemoglobin (totalHb) concentration changes as Δ oxyHb, Δ deoxygenated Hb, and Δ totalHb respectively in an arbitrary unit ($\mu\text{M cm}$) using the following equations:

$$\Delta\text{oxyHb} = -1.4887 \times \Delta\text{abs}_{780} + 0.5970 \times \Delta\text{abs}_{805} + 1.4847 \times \Delta\text{abs}_{830}$$

$$\Delta\text{deoxygenated Hb} = 1.8545 \times \Delta\text{abs}_{780} + (-0.2394) \times \Delta\text{abs}_{805} + (-1.0947) \times \Delta\text{abs}_{830}$$

$$\Delta\text{totalHb} = \Delta\text{oxyHb} + \Delta\text{deoxygenated Hb},$$

where Δabs indicated changes in light absorption at the corresponding wavelength. Since concurrent fNIRS–fMRI studies suggest superior signal to noise ratio and sensitivity of Δ oxyHb (Cui et al., 2011; Strangman et al., 2002), we mainly focused on our analysis using Δ oxy-Hb data but we also performed the same statistical analysis with Δ deoxygenated-Hb data. Raw data of hemodynamic signals from individual channels was low-pass-filtered through a 25-point Savitzky–Golay filter (Savitzky and Golay, 1964) and averaged. Baseline correction was applied to the averaged data so that the value at the task onset was set to zero. The amplitude of hemodynamic signal was further normalized by dividing the averaged values by the standard deviation of those during the 10 s before the task onset. The waveforms of hemodynamic signal were manually inspected and those from two subjects were contaminated with severe motion artifacts in their second trial of the WM and NM conditions. We incorporated hemodynamic signal only from the first trial for further analysis for those subjects.

The coordinates of the head landmarks and probe positions were used to estimate the position of each channel in the Montreal Neurological Institute standard brain space (Okamoto and Dan, 2005; Ye et al., 2009). Regions of interest (ROIs) were set to the FPC and the MTG. With each ROI, we chose the channels whose registration probability was more than 80%, calculated the grand mean waveform over the four game runs or trials, and determined the representative channel with maximum mean- Δ oxyHb amplitude. Therefore, we picked a single channel from each ROI and the channel was constant for the two conditions within each subject.

Subjects were divided into three groups based on performance accuracy using cluster analysis with unweighted pair group method with arithmetic mean (Sokal and Michener, 1958): high, intermediate, and low to examine the signal waveform. Performance accuracy was determined with respect to mean number of temporally accurate steps throughout the experiment, which corresponded to more than 50, 20–50, and 0–20% of temporally accurate steps for high, intermediate, and low performance groups, respectively. We then performed two types of analyses with hemodynamic signals in the FPC and MTG. First, we determined the time to the positive and negative peak of the Δ oxyHb and Δ deoxygenated Hb signals, respectively, between 10 s prior and 10 s after the task period at each ROI in each condition (the “peak time” hereinafter). Second, we calculated the integral of the waveform of normalized hemodynamic signal during the task period at each ROI in each condition (area under the curve: “AUC” hereinafter). We further calculated the difference of normalized hemodynamic signals between the two conditions by subtracting those in the NM condition from those in the WM condition (WM–NM) and their AUC at each ROI. The AUC(WM–NM) was calculated from the mean WM and NM waveforms that were averaged across two trials and therefore a single value was given to a single subject in each ROI. Since the number of accurate steps was comparable between two trials of the same condition within subjects ($P > 0.05$; Wilcoxon signed-rank test), all values were averaged between the two trials of the same condition for further statistical analysis. We compared the number of accurate steps in the two conditions using paired t-tests.

In order to determine the functional relationship between the accuracy of motor behavior and the temporal and cumulative characteristics of the hemodynamic activity of the FPC and MTG, we applied correlation analysis to these parameters. Either Pearson's correlation coefficient or Spearman's rank correlation coefficient was calculated depending on the result of the Shapiro–Wilk normality test. We considered P values <0.05 to be statistically significant.

Results

Raw data indicating region- and performance-specific oxy-hemoglobin responses during dance video gameplay

Fig. 2a depicts a representative result of projected optical channels on the cortical surface; with Figs. 2b and c depicting typical activation patterns for high- and low-performance players during the WM (red trace) and NM (blue trace) conditions, respectively. The activation in the larger measured area had two foci, namely the MTG and the FPC, which were defined as ROIs in our experiment and are indicated in Figs. 2b and c with bold boxes.

Figs. 3a–d indicates the complex interactions between behavioral performance, ROI and influence of music on timing and amplitude of signal. All subjects showed a bell-shaped Δ oxyHb waveform in the MTG in both WM and NM gameplay (Figs. 3b and d). The time to peak amplitude for high- and low-performance as well as the presence of music appeared to influence the characteristics of the resultant Δ oxyHb waveform (detailed analysis below). In both WM and NM, the signal peaked earliest in the low-performance group, followed by the intermediate and then high-performance players. All signals returned to baseline at the end of the rest period.

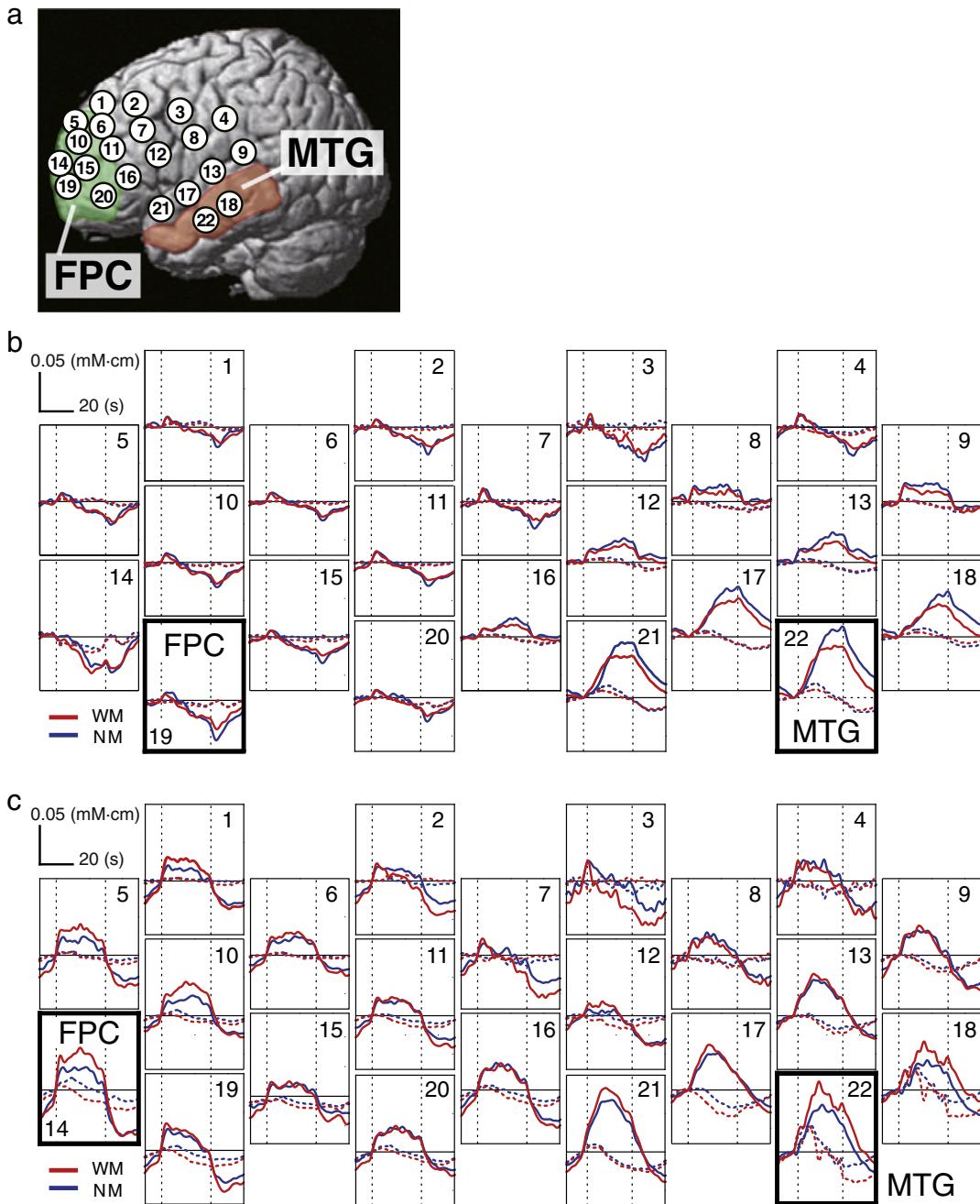


Fig. 2. Representative hemodynamic responses for high- and low-performance players. a. A representative result of projection of the optical channels onto the cortical surface. ROIs were set to the frontopolar cortex (FPC) and the middle temporal gyrus (MTG). b. Time-courses of ΔoxyHb (solid line) and $\Delta\text{deoxygenated Hb}$ (dotted line) responses of a high-performance player from two single trials of the “with music (WM)” and “no music (NM)” conditions. The number of temporally accurate step was 151 and 161 in the WM and NM conditions, respectively. c. Time-courses of ΔoxyHb (solid line) and $\Delta\text{deoxygenated Hb}$ (dotted line) responses of a low-performance player from two single trials of the WM and NM conditions. The number of temporally accurate step was 1 and 18 in the WM and NM conditions, respectively. Bold boxes indicate the channels corresponding to the two ROIs of the FPC and the MTG.

In contrast, ΔoxyHb waveforms in the FPC were strikingly different among players based on performance. High-performance players typically showed a small early ΔoxyHb increase followed by a rapid negative decline in ΔoxyHb during the task period (Figs. 3a and c, red trace), while low-performance players showed a similar early increase in ΔoxyHb , but the overall shape of the response was box-car shaped with prolonged activation throughout the duration of the task (Figs. 3a and c, blue trace). The negative signal of the high-performance players displayed a second peak at around 32 s and then returned to baseline at the end of the rest period. The positive boxcar signal of the low performance players ended at about 30 s, the end of the task period, and

returned to baseline by the end of the rest period. The presence or absence of music appeared to influence the response of all subjects in the FPC minimally.

Correlation among peak time, AUC, and temporal accuracy of dance steps

The consistent bell-shaped response in the MTG from all subjects (Figs. 3b and d) allowed us to compare duration of time to maximal signal. There was a significant correlation between the time to maximum amplitude of signal (peak time) of ΔoxyHb in the MTG and number of accurate steps in both the WM and NM conditions

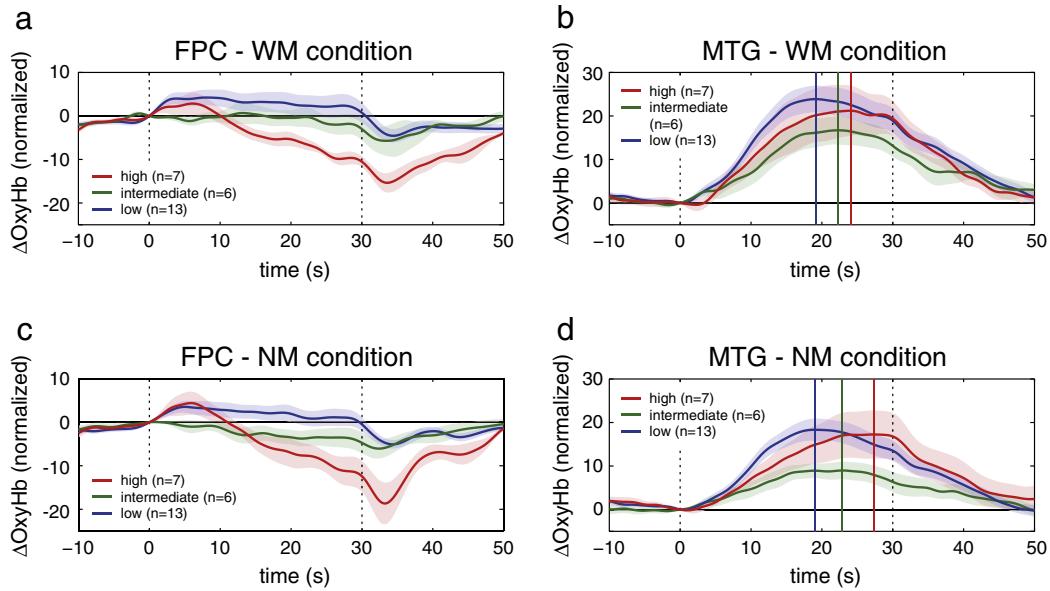


Fig. 3. Comparison of mean ΔOxyHb time-course responses in players of different performance groups. All subjects were divided into high- (red), intermediate- (green), and low- (blue) performance groups, corresponding to those who made more than 50, 20–50, and 0–20% of temporally accurate steps among the total number of steps, respectively. a and b. Comparison of mean ΔOxyHb time-course responses in players of different performance groups in the WM condition in the FPC (a) and in the MTG (b). c and d. Comparison of mean ΔOxyHb time-course responses in players of different performance groups in the NM condition in the FPC (c) and in the MTG (d). Dotted vertical lines indicate the 30 s task period. The horizontal line indicates the baseline. Standard error of the mean was represented as a shaded area around the waveform in corresponding colors. The solid vertical line in b and d indicate the peak time of each of the three subgroups in corresponding colors.

(Spearman's rank correlation coefficient $R = 0.562$, $P < 0.03$ and $R = 0.573$, $P < 0.02$, respectively; Figs. 4a–b). There was no significant correlation between AUC and the number of accurate steps (performance) in the MTG in either condition ($P > 0.1$).

A similar temporal analysis was not feasible with ΔOxyHb response in the FPC because of the skill-based differences in the shape of the response waveforms (Figs. 3a and c). The FPC also showed a significant negative correlation between the number of accurate steps and AUC in both the WM and NM conditions ($R = -0.485$, $P < 0.02$ and $R = -0.518$, $P < 0.01$, respectively), that is, the lower the number of accurate steps, the higher the FPC activation (Figs. 5a–b).

Finally we determined the correlation between MTG and FPC activations and found that there was a significant negative correlation between the peak time of ΔOxyHb at the MTG and AUC in the FPC (Pearson's correlation coefficient $R = -0.403$, $P < 0.05$; Fig. 6) in the NM condition. This indicates that players who showed sustained

activation (longer peak time) in the MTG tended to show lower activation (or deactivation) in the FPC while players who showed short-lived activation at the MTG tended to show higher and box-car-shaped activation in the FPC.

We also analyzed the time course of $\Delta\text{deoxygenatedHb}$ responses in the two ROIs. As shown in Figs. 2b and c, all subjects showed decreasing $\Delta\text{deoxygenatedHb}$ responses in the MTG, following the bell-shaped ΔOxyHb waveform in both WM and NM conditions. However, we did not find any statistically significant correlation between the negative peak time of $\Delta\text{deoxygenatedHb}$ signal and the number of temporally accurate steps. The $\Delta\text{deoxygenatedHb}$ response in the FPC also showed negative deflection (Figs. 2b and c) despite the performance-dependent variety in the shape of ΔOxyHb responses. The negative peak time of $\Delta\text{deoxygenatedHb}$ response was negatively correlated with the number of accurate steps in WM condition (Spearman's rank correlation coefficient $R = -0.521$, $P < 0.01$). This corresponds to the short-lasting and prolonged positive

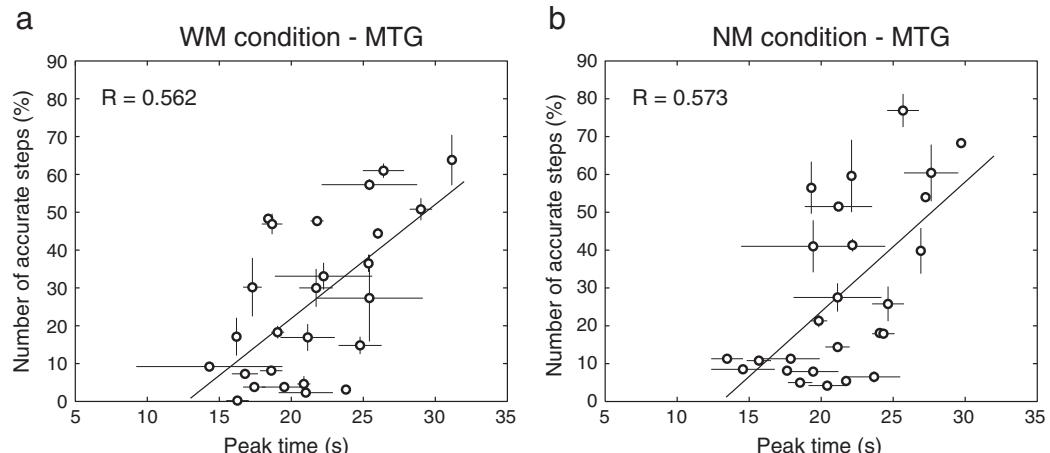


Fig. 4. Correlation between peak time in the MTG and the number of accurate steps in the WM condition (a) and in the NM condition (b). Each dot represents a single subject. Error bars show standard error of the mean from two trials. Statistically significant correlation was found in both conditions.

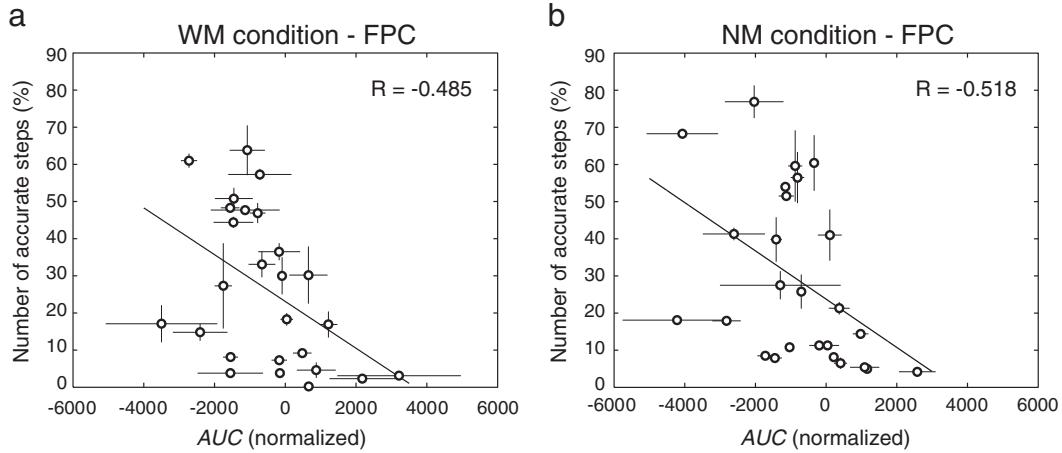


Fig. 5. Correlation between AUC in the FPC and the number of accurate steps in the WM condition (a) and in the NM condition (b). Each dot represents a single subject. Error bars show standard error of the mean from two trials. Statistically significant correlation was found in both conditions.

Δ oxyHb peak observed in the high-performance and low-performance players, respectively (Figs. 2b and c and 3a and c).

Behavior and effect of eliminating auditory cues

Anecdotally, when players interact with rhythm or music games like Dance Dance Revolution, they often state that the game is harder to interact with in the absence of rhythm. To test the hypothesis that performance improves in the presence of music, we determined differences in behavioral outcomes with respect to playing the game with and without music. Results indicate that the number of accurate steps were not significantly different in the NM condition ($29.0 \pm 4.5\%$ of total steps) compared to the WM condition ($26.4 \pm 4.1\%$ of total steps) (Wilcoxon signed-rank test, $P = 0.062$).

While there were no behavioral differences between the NM and WM conditions, we further investigated the conditional changes of neuronal activation in each ROI during WM or NM conditions. We found a significant negative correlation between AUC(WM–NM) of Δ oxyHb response and the number of accurate steps at the MTG (Spearman's rank correlation coefficient $R = -0.518$, $P < 0.01$; Fig. 7a). This indicates that elimination of auditory cues increased cumulative Δ oxyHb response at the MTG in high-performance players, while it suppressed that in low-performance players. We also found a significant positive correlation between AUC(WM–NM) of Δ deoxyHb response and the number of accurate steps at the MTG (Spearman's rank correlation coefficient $R = 0.504$, $P < 0.01$), which supports the

increased deflection of hemodynamic response in the NM condition at the MTG in high-performance players.

The same correlation was not found in the FPC ($P > 0.1$; Fig. 7b). However, it is interesting that high-performance players showed similar FPC activation between the WM and NM conditions (nearly 0 in WM–NM), while low-performance players showed either positive or negative WM–NM differences. To help understand why low-performance players displayed these varying results we employed a post-experiment questionnaire which differentiated between the players who reported that the WM condition required more attentional effort than the NM condition from players who reported the opposite. There was a significant difference in the mean AUC(WM–NM) between the subjects who required more attentional effort in the WM condition (614.6 ± 308.6) and in the NM condition (-379.0 ± 218.0) compared to their counterparts (Mann–Whitney U-test, $P < 0.01$; Fig. 7b). This suggests that the FPC activation was increased in the condition where the subjects reported greater attentional effort. To cancel out the subject's preference, we also calculated the absolute value of AUC(WM–NM) ($|AUC(WM–NM)|$). We found a significant correlation between $|AUC(WM–NM)|$ and the number of accurate steps (Spearman's rank correlation coefficient $R = -0.664$, $P < 0.01$; Fig. 7c), indicating that low-performance players tended to show different FPC activations between the WM and NM conditions while high-performance players showed similar activation in both conditions. Neither AUC(WM–NM) nor $|AUC(WM–NM)|$ of Δ deoxyHb signal in the FPC showed statistically significant correlation with the number of accurate steps.

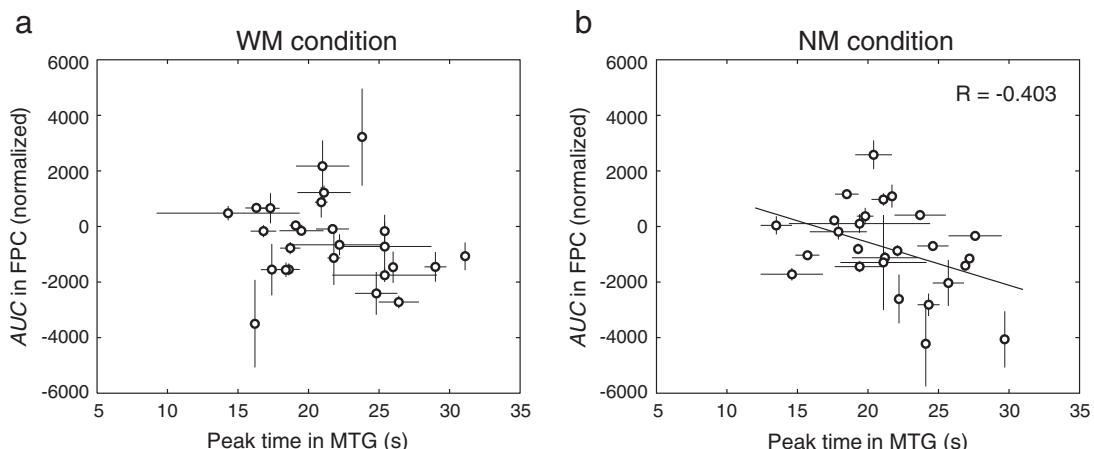


Fig. 6. Correlation between peak time in the MTG and AUC in the FPC. Each dot represents a single subject. Error bars show standard error of the mean from two trials. Statistically significant correlation was found in the NM condition but not in the WM condition.

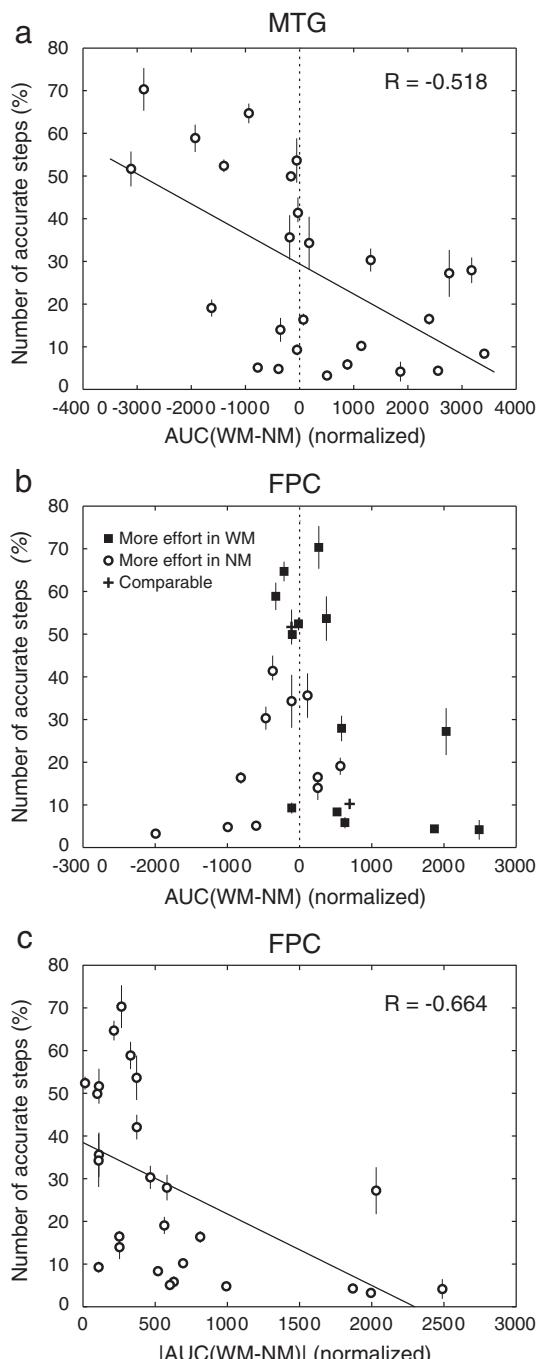


Fig. 7. Correlation between AUC(WM-NM) and the number of accurate steps in the MTG (a) and in the FPC (b, c). a. AUC(WM-NM) negatively correlates with the number of accurate steps in the MTG. Each dot represents a single subject. b. AUC(WM-NM) in the FPC shows a large variance in the low-performance subjects. Subjects were divided into three groups according to the response to the post-experimental attentional effort questionnaire. Filled boxes indicate subjects who responded that they spent more attentional effort in the WM condition. Open circles indicate subjects who responded that they spent more attentional effort in the NM condition. Two subjects responded that both conditions equally required attentional effort and they were represented by cross marks. c. Absolute value of AUC(WM-NM) in the FPC negatively correlates with the number of accurate steps in the FPC. Error bars in the vertical axis show standard error of the mean from four trials.

Discussion

We used a dance simulation video game to study how visual and rhythmic auditory cues are processed in the cortical areas of multimodal integration and decision making to achieve coordinated motor outputs.

A rapid increase in FPC signal suggests the enhanced attention at task onset. This was compared to gradual onset of MTG activity with greater peak amplitude indicating the ongoing sensory integration occurring throughout the task. Our study demonstrated that the temporal accuracy of dance steps positively correlated with the persistent activation of the MTG, while it negatively correlated with the cumulative suppression of the FPC. Elimination of auditory cues affected these two ROIs in a performance-dependent fashion. Within-subject comparison of hemodynamic responses demonstrated that high-performance players showed an increase but low-performance players displayed a decrease of cumulative amount of Δ OxyHb in the NM condition compared to that in the WM condition in the MTG. On the other hand, high-performance players showed relatively small variance in FPC activity regardless of the presence of auditory cues, while low-performance players showed larger differences between two conditions. Our results suggest that the time-course and cumulative characteristics of Δ OxyHb responses obtained from fNIRS in these association cortical sites could be good predictors of the temporal accuracy of motor responses.

MTG as a bottom-up integration center for multimodal auditory and visual cues

The MTG interposes auditory and visual processing streams and therefore has been considered as a higher integration site of auditory and visual information (Beauchamp et al., 2004; Calvert et al., 2001; Fuhrmann Alpert et al., 2008; Visser et al., 2012) to execute an appropriate movement response in the right timing (Fridman et al., 2006; Tankus and Fried, 2012). The persistent activation of the MTG is crucial to perform the current experimental task since dance video gameplay requires continuous integration of visual (the arrows moving upwards and the game feedback to the selected motor response) and auditory (background music) cues to generate appropriate motor output.

Considering that the increase of Δ OxyHb amplitude represents a persistent consumption of oxygen related to the continuous neuronal activation, we suggest that the time to peak signal correlates with the stability of neuronal processing in the corresponding cortical area. Our results clearly demonstrate the positive relationship between the stability of the MTG activation and the temporal accuracy of behavioral responses (Fig. 4). The prolonged time to peak of Δ OxyHb responses in the MTG suggests continuous processing of visual and auditory integration during almost the entire duration of task period, contributing to the increased number of temporally accurate steps (Fig. 4a). Likewise, the shorter time to peak seen in players with a lower number of temporally accurate steps suggests a cessation of, or the intermittent dropout from, the continuous integration of visual and auditory cues in the MTG during the task period. This is likely due to a limited capacity of multimodal integration. Such subjects may occasionally fail to bind visual and auditory cues for controlling the motor output resulting in a lower number of accurate steps.

Interestingly, this relationship between the time to peak and the number of accurate steps was preserved even in the NM condition, in which auditory cues were not provided (Fig. 4b). There is intriguing evidence that crossmodal recruitment of primary sensory cortices in the presence of only unimodal cues can occur in the complete absence of any explicit conditioning, semantic factors, or long-term association (Zangenehpour and Zatorre, 2010). The single exposure of the song to subjects during practice could be explained by these results. Since the current task of the dance video game requires highly rhythmic movements even without the auditory cues, subjects are likely to internally generate aspects of the rhythm (Grahn, 2009) and integrate it with the visual cues. Indeed, most of the high-performance players reported in the post-experiment interview that they could spontaneously imagine the rhythm of the game song even without the background music. Our results suggest that the sustained activation of the MTG indicates the continuous integration of the given visual cues with internally-generated rhythm, and that subjects who could persistently integrate

the multimodal information achieved an increased number of accurate steps as in the WM condition.

FPC as a top-down regulation center for coordination of multiple cognitive processes

Although we observed typical bell-shaped Δ OxyHb responses in the MTG in all subjects, the time-course of Δ OxyHb and directionality of responses in the FPC varied across subjects (Figs. 2 and 3). The AUC analysis (Fig. 5) further highlights the difference of Δ OxyHb responses between high- and low-performance players, indicating that suppressed Δ OxyHb responses in the FPC correlated with the increased number of temporally accurate steps. Because of its unique feature of lower density of cell bodies and higher number of dendritic spines per cell and spine density, the FPC has been considered a supramodal regulation area, especially in cases in which outcomes of two or more separate cognitive operations need to be integrated to achieve a higher behavioral goal (Ramnani and Owen, 2004). A number of studies reported that the FPC plays a crucial role in “predictive coding” (Summerfield et al., 2006; Chambon et al., 2011) or “prospective memory” (Burgess et al., 2001), functioning to hold the intention of future behavior until the right timing for execution based on the environmental feedback. The dance video game used in this study undeniably required this prospective memory; subjects had to prepare the correct sequence of the steps from the rhythmic and visual cues, keep them in mind for seconds until the visual cues scrolled up to the response area, and then send the appropriate motor command to step onto the correct arrow-button.

The difference in the cumulative activity of Δ OxyHb responses in the FPC among subjects suggests the required cognitive load to perform the task. Similarly, in an fNIRS study comparing prefrontal cortex activity in surgeons and medical students while they performed a knot-tying task, increased Δ OxyHb was measured in the less experienced subjects (Leff et al., 2008). Together with the results in the MTG, the Δ OxyHb responses in the FPC with fast and small positive peak followed by persistent suppression in the high-performance players (red line in Figs. 3a and c) suggest that they required the prospective memory at the beginning of the task (such as finding the rhythm for their first step in the task period) but no longer required it afterwards since they could continuously and successfully integrate visual and rhythmic cues in the MTG to make accurate steps. Alternatively, the boxcar-shaped Δ OxyHb responses in the low-performance players (blue line in Figs. 3a and c) suggest a persistent cognitive load in the FPC for most of the task period, likely to compensate the insufficient integrative ability of visual and rhythmic cues in the MTG to continuously generate the required motor output. Indeed, the peak time in the MTG negatively correlated with the AUC in the FPC in the NM condition, suggesting a complementary role of these two multimodal integration sites in this task (Fig. 6). FPC may be particularly important in this multisensory integration process only when the MTG fails to integrate the multisensory information.

Subjects of intermediate-performance group displayed a fluctuated shape of Δ OxyHb responses compared to those of the other two groups in the FPC. Some showed increase with slow latency, some showed gradual decrease without rapid positive peak and others showed bipolar shaped responses (data not shown). This resulted in the cancellation of an early positive peak at the beginning of the task period in the averaged response (green line in Figs. 3a and c), while the negative deflection at the end of the task period lies in between the red high-performance and blue low-performance signals. These fluctuated initial response shapes may represent the transitional state from the low- to high-performance.

Performance-dependent responses in the bottom-up and top-down integration centers during multimodal vs. unimodal cued motor task

Comparison between the WM and NM conditions demonstrated the different degree of dependency upon the auditory cue in high-

and low-performance players. The MTG activity was enhanced in high-performance players, while suppressed in low-performance players when no music cue was provided (Fig. 7a). The enhanced MTG activity in the high-performance players with the NM condition suggests additional processing to integrate the visual and the internally-generated rhythm to produce accurate steps. Indeed, the cumulative Δ OxyHb responses in the FPC showed little change between the WM and NM conditions in the high-performance players (Figs. 7b, and c), confirming that their strategy for playing this dance video game – integrating visual and auditory cues – was consistent regardless of the source of the auditory cue (external or internal).

On the other hand, low-performance players showed suppressed MTG activity in the NM condition (Fig. 4a), suggesting less successful multimodal integration. These results raise a hypothesis that the low-performance player used different strategies between the two conditions. They used both visual and auditory cues and tried to integrate them when those were provided in the WM condition; while in the NM condition they mostly relied on the visual information to make steps. Therefore the cognitive load for multimodal integration in the MTG was decreased in the NM condition. The large variance in AUC(WM-NM) in the FPC further supports this hypothesis (Figs. 7b and c). The low-performance subjects who reported that the WM condition was more challenging compared to the NM condition demonstrated enhanced FPC activation in the WM condition than in the NM condition, and vice versa. This result indicates that when the low-performance subjects had a preference for one condition, the cognitive load in the FPC was increased in the other condition. Since this dance video game requires persistent integration of audiovisual cues with internally generated rhythms for motor output, the result suggests that relying on a single modality cue is insufficient to achieve a large number of accurate steps and therefore multimodal integration in the MTG is crucial.

Our findings of similar temporal stepping accuracy during WM and NM gameplay performance contradict previous findings of improved performance with combined audio-visual stimuli compared to a single modality (audio or visual) alone (Werner and Noppeney, 2010). While some experimental paradigms suggest that vision is the dominant sensory modality as seen in the Colavita effect (Spence, 2009); comparisons of visual versus auditory beat perception alone found that subjects were more accurate in synchronizing movement with an auditory modality (Jantzen et al., 2005; Patel et al., 2005). These conflicting results may be explained by the reported differences found in individuals in auditory beat perception (Grahn and McAuley, 2009; Repp, 2005). Individuals with greater sensitivity to changes in rhythm manipulated to slow down or speed up were found to have differences in cortical activity compared to less sensitive individuals. However, this research did not investigate the link between rhythm perception and accuracy of motor output synchronization.

Limitations

The interpretation of our findings could be limited by several factors. First, we did not limit our subjects to rank novice or highly experienced dance videogame players. Rather, we chose to have a representation of playing ability across the spectrum in order to explore the interaction of accuracy on neural substrate activity. All subjects interacted with an identical game-song environment to attempt to control for environmental variables. Future investigations will examine changes in neural activity with training, using novice players as well as determining differences in more complex environments. Second, we did not study a condition with only auditory input as it is not possible to play this game without the visual arrows indicating time to step. Third, use of visual stimuli requires recruitment of cortical regions that are distinct from those of auditory (Witt et al., 2008), however, we were limited to the size of the fNIRS optode array. Fourth, the fixed inter-optode distance in our NIRS system may have contaminated cerebral responses

with extracerebral hemodynamic changes, although the system is optimally designed to record cortical hemodynamic responses.

Possible application of fNIRS temporal dynamics on rehabilitation and motor learning

A particular interest for future medical application of this study is that the well-trained subjects (frequent players of this game) can integrate visual cues with their own internally generated rhythms as if there were auditory rhythmic cues available even in the situation without background music. Considering that rhythmic auditory cueing in the form of music or a metronome is often used to help patients perform stepping tasks for Parkinson's disease patients who suffer from 'freezing' (de Bruin et al., 2010; Rochester et al., 2010), training such patients with a dance video game may help them to robustly generate spontaneous rhythmic cues for their intended movement even when external auditory cues are not provided in their daily life.

Conclusions

We studied the cortical responses in the multimodal integration sites to understand the process of temporally accurate motor output based on the audiovisual cues. We propose that changes in these relationships can be monitored to gage performance changes in motor learning. Differences in oxy-hemoglobin response with performance in the FPC confirmed the regulatory role of this cortical region and suggest a novel mechanism of motor learning. The temporal persistency of oxy-hemoglobin signals in the MTG correlates with the accuracy of the behavioral output. This suggests that continuous integration of audiovisual cues in the MTG plays an important role to maintain temporal accuracy of the motor output. Our results further demonstrate the unique feature of fNIRS; temporal characteristics of the hemodynamic signals such as peak time in a particular cortical area contain meaningful information on the dynamics of neuronal activity, predicting the accuracy of behavioral output. Since such characteristics of the hemodynamic signals are difficult to obtain in the conventional fMRI experiment, the advantage of fNIRS in measuring hemodynamic changes with higher temporal resolution should be emphasized and combined with the findings from other neuroimaging modalities to further explore the mechanisms of neuronal systems.

Acknowledgments

The authors wish to thank the Grants-in-Aid for Scientific Research (from the University of Education, Science, and Culture of Japan), grant 23700632, and a Research Grant from Hayao Nakayama Foundation for Science & Technology and Culture.

Conflict of interest

The authors have no conflict of interest to disclose.

References

- Bangert, M., Peschel, T., Schlaug, G., Rotte, M., Drescher, D., Hinrichs, H., Altenmüller, E., 2006. Shared networks for auditory and motor processing in professional pianists: evidence from fMRI conjunction. *Neuroimage* 30 (3), 917–926.
- Baumann, S., Koeneke, S., Schmidt, C.F., Meyer, M., Lutz, K., Jäncke, L., 2007. A network for audio-motor coordination in skilled pianists and non-musicians. *Brain Res.* 1161, 65–78.
- Beauchamp, M.S., Lee, K.E., Argall, B.D., Martin, A., 2004. Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron* 41 (5), 809–823. [http://dx.doi.org/10.1016/S0896-6273\(04\)00070-4](http://dx.doi.org/10.1016/S0896-6273(04)00070-4) (pii).
- Burgess, P.W., Quayle, A., Frith, C.D., 2001. Brain regions involved in prospective memory as determined by positron emission tomography. *Neuropsychologia* 39 (6), 545–555.
- Calvert, G.A., Hansen, P.C., Iversen, S.D., Brammer, M.J., 2001. Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *Neuroimage* 14 (2), 427–438. <http://dx.doi.org/10.1006/nimg.2001.0812>.
- Chambon, V., Domenech, P., Pacherie, E., Koechlin, E., Baraduc, P., Farrer, C., 2011. What are they up to? The role of sensory evidence and prior knowledge in action understanding. *PloS one* 6 (2), e17133.
- Chatrani, G.E., Lettich, E., Nelson, P.L., 1985. Ten percent electrode system for topographic studies of spontaneous and evoked EEG activity. *Am J. EEG Technol.* 25, 83–92.
- Cope, M., Delpy, D.T., Reynolds, E.O., Wray, S., Wyatt, J., van der Zee, P., 1988. Methods of quantitating cerebral near infrared spectroscopy data. *Adv. Exp. Med. Biol.* 222, 183–189.
- Cui, X., Bray, S., Bryant, D.M., Glover, G.H., Reiss, A.L., 2011. A quantitative comparison of NIRS and fMRI across multiple cognitive tasks. *Neuroimage* 54 (4), 2808–2821. <http://dx.doi.org/10.1016/j.neuroimage.2010.10.069>.
- de Bruin, N., Doan, J.B., Turnbull, G., Suchowersky, O., Bonfield, S., Hu, B., Brown, L.A., 2010. Walking with music is a safe and viable tool for gait training in Parkinson's disease: the effect of a 13-week feasibility study on single and dual task walking. *Park. Dis.* 483530. <http://dx.doi.org/10.4061/2010/483530>.
- Ferrari, M., Quaresima, V., 2012. A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application. *Neuroimage* 63 (2), 921–935. <http://dx.doi.org/10.1016/j.neuroimage.2012.03.049>.
- Ferrari, M., Mottola, L., Quaresima, V., 2004. Principles, techniques, and limitations of near infrared spectroscopy. *Can. J. Appl. Physiol.* 29 (4), 463–487. <http://dx.doi.org/10.1139/h04-031>.
- Fridman, E.A., Immisch, I., Hanakawa, T., Bohlhalter, S., Waldvogel, D., Kansaku, K., Hallett, M., 2006. The role of the dorsal stream for gesture production. *Neuroimage* 29 (2), 417–428. <http://dx.doi.org/10.1016/j.neuroimage.2005.07.026>.
- Fuhrmann Alpert, G., Hein, G., Tsai, N., Naumer, M.J., Knight, R.T., 2008. Temporal characteristics of audiovisual information processing. *J. Neurosci.* 28 (20), 5344–5349. <http://dx.doi.org/10.1523/JNEUROSCI.5039-07.2008>.
- Grahn, J.A., 2009. The role of the basal ganglia in beat perception: neuroimaging and neuropsychological investigations. *Ann. N. Y. Acad. Sci.* 1169, 35–45. <http://dx.doi.org/10.1111/j.1749-6632.2009.04553.x> (NYAS04553 [pii]).
- Grahn, J.A., McAuley, J.D., 2009. Neural bases of individual differences in beat perception. *Neuroimage* 47 (4), 1894–1903. <http://dx.doi.org/10.1016/j.neuroimage.2009.04.039> (S1053-8119(09)00387-5 [pii]).
- Hasegawa, T., Matsuki, K.I., Ueno, T., Maeda, Y., Matsue, Y., Konishi, Y., Sadato, N., 2004. Learned audio-visual cross-modal associations in observed piano playing activate the left planum temporale. An fMRI study. *Cogn. Brain Res.* 20 (3), 510–518.
- Hoshi, Y., 2003. Functional near-infrared optical imaging: utility and limitations in human brain mapping. *Psychophysiology* 40 (4), 511–520. <http://dx.doi.org/10.1111/1469-8986.00053>.
- Jäncke, L., Loose, R., Lutz, K., Specht, K., Shah, N.J., 2000. Cortical activations during paced finger-tapping applying visual and auditory pacing stimuli. *Cogn. Brain Res.* 10 (1), 51–66.
- Jantzen, K.J., Steinberg, F.L., Kelso, J.A., 2005. Functional MRI reveals the existence of modality and coordination-dependent timing networks. *Neuroimage* 25 (4), 1031–1042. <http://dx.doi.org/10.1016/j.neuroimage.2004.12.029>.
- Leff, D.R., Elwell, C.E., Orihuela-Espina, F., Atallah, L., Delpy, D.T., Darzi, A.W., Yang, G.Z., 2008. Changes in prefrontal cortical behaviour depend upon familiarity on a bimodal co-ordination task: an fNIRS study. *Neuroimage* 39 (2), 805–813. <http://dx.doi.org/10.1016/j.neuroimage.2007.09.032>.
- Okamoto, M., Dan, I., 2005. Automated cortical projection of head-surface locations for transcranial functional brain mapping. *Neuroimage* 26 (1), 18–28. <http://dx.doi.org/10.1016/j.neuroimage.2005.01.018> (S1053-8119(05)00050-9).
- Patel, A.D., Iversen, J.R., Chen, Y., Repp, B.H., 2005. The influence of metricality and modality on synchronization with a beat. *Exp. Brain Res.* 163 (2), 226–238. <http://dx.doi.org/10.1007/s00221-004-2159-8>.
- Ramnani, N., Owen, A.M., 2004. Anterior prefrontal cortex: insights into function from anatomy and neuroimaging. *Nat. Rev. Neurosci.* 5 (3), 184–194. <http://dx.doi.org/10.1038/nrn1343>.
- Repp, B.H., 2005. Sensorimotor synchronization: a review of the tapping literature. *Psychon. Bull. Rev.* 12 (6), 969–992.
- Rochester, L., Rafferty, D., Dotchin, C., Msuya, O., Minde, V., Walker, R.W., 2010. The effect of cueing therapy on single and dual-task gait in a drug naive population of people with Parkinson's disease in northern Tanzania. *Mov. Disord.* 25 (7), 906–911. <http://dx.doi.org/10.1002/mds.22978>.
- Sakai, K., Ramnani, N., Passingham, R.E., 2002. Learning of sequences of finger movements and timing: frontal lobe and action-oriented representation. *J. Neurophysiol.* 88 (4), 2035–2046.
- Savitzky, A., Golay, M.J.E., 1964. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* 36, 1627–1639.
- Sokal, R., Michener, C., 1958. A statistical method for evaluating systematic relationships. *Univ. Kans. Sci. Bull.* 38, 1409–1438.
- Spence, C., 2009. Explaining the Colavita visual dominance effect. *Prog. Brain Res.* 176, 245–258. [http://dx.doi.org/10.1016/S0079-6123\(09\)17615-X](http://dx.doi.org/10.1016/S0079-6123(09)17615-X).
- Strangman, G., Culver, J.P., Thompson, J.H., Boas, D.A., 2002. A quantitative comparison of simultaneous BOLD fMRI and NIRS recordings during functional brain activation. *Neuroimage* 17 (2), 719–731.
- Summerfield, C., Egner, T., Greene, M., Koechlin, E., Mangels, J., Hirsch, J., 2006. Predictive codes for forthcoming perception in the frontal cortex. *Science* 314 (5803), 1311–1314.
- Tachibana, A., Noah, J.A., Bronner, S., Ono, Y., Onozuka, M., 2011. Parietal and temporal activity during a multimodal dance video game: an fNIRS study. *Neurosci. Lett.* 503 (2), 125–130. <http://dx.doi.org/10.1016/j.neulet.2011.08.023>.
- Tankus, A., Fried, I., 2012. Visuomotor coordination and motor representation by human temporal lobe neurons. *J. Cogn. Neurosci.* 24 (3), 600–610. http://dx.doi.org/10.1162/jocn_a_00160.
- Visser, M., Jefferies, E., Embleton, K.V., Lambon Ralph, M.A., 2012. Both the middle temporal gyrus and the ventral anterior temporal area are crucial for multimodal semantic

- processing: distortion-corrected fMRI evidence for a double gradient of information convergence in the temporal lobes. *J. Cogn. Neurosci.* 24 (8), 1766–1778. http://dx.doi.org/10.1162/jocn_a_00244.
- Werner, S., Noppeney, U., 2010. Superadditive responses in superior temporal sulcus predict audiovisual benefits in object categorization. *Cereb. Cortex* 20 (8), 1829–1842. <http://dx.doi.org/10.1093/cercor/bhp248>.
- Witt, S.T., Laird, A.R., Meyerand, M.E., 2008. Functional neuroimaging correlates of finger-tapping task variations: an ALE meta-analysis. *Neuroimage* 42 (1), 343–356. <http://dx.doi.org/10.1016/j.neuroimage.2008.04.025>.
- Ye, J.C., Tak, S., Jang, K.E., Jung, J., Jang, J., 2009. NIRS-SPM: statistical parametric mapping for near-infrared spectroscopy. *Neuroimage* 44 (2), 428–447. <http://dx.doi.org/10.1016/j.neuroimage.2008.08.036>.
- Zangenehpour, S., Zatorre, R.J., 2010. Crossmodal recruitment of primary visual cortex following brief exposure to bimodal audiovisual stimuli. *Neuropsychologia* 48 (2), 591–600. <http://dx.doi.org/10.1016/j.neuropsychologia.2009.10.022>.
- Zatorre, R.J., Chen, J.L., Penhune, V.B., 2007. When the brain plays music: auditory-motor interactions in music perception and production. *Nat. Rev. Neurosci.* 8 (7), 547–558.