Social Cognitive and Affective Neuroscience Advance Access published October 7, 2013

doi:10.1093/scan/nst147 SCAN (2013) 1 of 6

Changing the mind? Not really—activity and connectivity in the caudate correlates with changes of choice

Takehito Ito,[1] Daw-An Wu,[2] Toshiyuki Marutani,[1] Manami Yamamoto,[1] Hidenori Suzuki,[3] Shinsuke Shimojo,[2] and Tetsuya Matsuda[1]

[1]Brain Science Institute, Tamagawa University, 6-1-1, Tamagawa Gakuen, Machida, Tokyo 194-8610, Japan, [2]Division of Biology/Option Representative, Computation and Neural Systems, California Institute of Technology, 139-74, Pasadena, CA 91125, USA, and [3]Department of Pharmacology, Nippon Medical School, 1-1-5, Sendagi, Bunkyo-ku, Tokyo 113-8602, Japan

Changes in preference are inherently subjective and internal psychological events. We have identified brain events that presage ultimate (rather than intervening) choices, and signal the finality of a choice. At the first exposure to a pair of faces, caudate activity reflected the face of final choice, even if an initial choice was different. Furthermore, the orbitofrontal cortex and hippocampus exhibited correlations only when the subject had made a choice that would not change.

Keywords: caudate; changing mind; face preference; fMRI; gaze manipulation

INTRODUCTION

Why do we sometimes change our mind?

In the last several decades, psychologists have learned a lot about the neural basis of decision making (Kable *et al.*, 2009; Pennartz *et al.*, 2009). However, the distinction between sustained and transient decisions and the possibility that the decision may be overturned later have been largely neglected. We know from daily experiences that when we have to make a choice/decision, we sometimes reach a firm decision, but at other times we change our mind. However, so far we have not been able to identify the neural basis or the temporal dynamics of decision change or 'changing the mind'. Changing mind regarding a preference is especially interesting, because of its exclusively subjective and internal nature. Recent studies have identified the so-called 'binary attractor model' as being behind the behavioral change process involving single decisions, but they have not revealed much about the internal process that leads to a change in an already-made decision and its neural correlates (Resulaj *et al.*, 2009; Krajbich *et al.*, 2010; Albantakis and Deco, 2011). It has been difficult in the laboratory to obtain meaningful behavioral and neural data about the mechanisms involved in 'changing the mind' phenomenon, for a variety of methodological and technical reasons, including the difficulty to generate frequent 'real' responses associated with a change of mind and to find a paradigm to systematically manipulate preference decisions. In this study, we aimed to reveal the temporal dynamics of the changing mind in a serial time task, especially in terms of its neural basis. To the best of our knowledge, this was the first study that employed this approach; there were little evidences that could reveal the neural basis or the temporal dynamics of the decision change; that is the changing mind. Especially, the changing mind of preference decisions is interesting due to its exclusively subjective, internal nature.

Received 24 April 2013; Accepted 26 August 2013

This work was supported by Tamagawa University global Center of Excellence (GCOE) program of the Japanese Ministry of Education, Culture, Sports, and Technology. SS has been supported by the Japanese Science and Technology Agency (JST) CREST program (Implicit Interpersonal Communication). TI has been supported by the Grant-in-Aid for Young Scientists (B) (No. 24730628).

Correspondence should be addressed to Tetsuya Matsuda, Brain Science Institute, Tamagawa University, 6-1-1, Tamagawa Gakuen, Machida, Tokyo 194-8610, Japan. E-mail: tetsuya@lab.tamagawa.ac.jp, and Shinsuke Shimojo, Division of Biology/Option Representative, Computation and Neural Systems, California Institute of Technology, 139-74, Pasadena, CA 91125, USA. E-mail: sshimojo@caltech.edu

According to previous studies, attractive faces activate reward-related neural circuits. In particular, the striatum [putamen and nucleus accumbens (NAcc)]; the orbitofrontal cortex (OFC) is involved in face attractiveness judgment (Kim *et al.*, 2007; Tsukiura and Cabeza, 2011; Mende *et al.*, 2012). A recent study showed that there is a significant correlation between face attractiveness and hippocampal activation. This too suggests that the OFC and the hippocampus may be important for the evaluation of face attractiveness (Tsukiura and Cabeza, 2011). Furthermore, these results indicate that the evaluation of facial attractiveness is not only based on reward-related, but also on memory-related neural circuits.

The rationale behind the current study is as follows: since the preference decision and, in particular, its change are subjective and internal in nature, it is hard to explain them in terms of stimulus-driven, deterministic mechanisms. Therefore, we sought an alternative explanation as well as some physiological evidence to substantiate our rationale. We focused on the following three hypotheses regarding the neural correlates of the 'changing the mind' phenomenon.

(1) Changing the mind is 'real', in the sense that the entire brain consistently reverses the dominance in its activity, starting from the most upstream process in the neural information-processing cascade, possibly because of sensory or attentive modulations.
(2) Changing the mind is 'not real'. Although reversal takes place at the behavioral (and thus the downstream decoding) levels, such as the motor cortices, possibly due to some noise, the upstream subcortical areas do not change the dominance or the relative strength of their neural activity.
(3) There is always competition between two potential choices in neural circuits or brain activities. Alternative neural circuits are competing and suppressing each other, until one reaches the threshold to generate a choice action. The 'changing the mind' phenomenon reflects fluctuations in such a neural competition (cf. 'Competition hypothesis', Reynolds and Desimone, 1999; Resulaj *et al.*, 2009).

Currently available data are consistent with Hypothesis 2 and, at least partly, with Hypothesis 3 as explained below. For example, Kim *et al.* (2007) have identified different neural components of the preference decision-making process in the time domain that indicate a signal transfer from the NAcc in the ventral striatum to the OFC

(Kim *et al.*, 2007). To test these hypotheses, we analyzed the temporal dynamics of the neural processes involved in face preference decision-making. To obtain frequent occurrences of mind changing in a systematically manipulated way, we employed the gaze manipulation paradigm (Shimojo *et al.*, 2003).

Unlike previous related studies (e.g. Tsukiura and Cabeza, 2011), which employed a task of absolute attractiveness evaluation of a single face, participants in the current study performed a two-alternative forced-choice task twice. We examined differences and fluctuations in neural activity between the choice-changed cases and choice-not-changed cases. The results showed that, if the caudate showed a higher activation to a face at first sight, the participant chose that face at the time of the second decision. This is true regardless of whether the face with the higher caudate activation happened to be the first choice or not. Moreover, the OFC and the hippocampus exhibited a functional correlation at the first decision period. In contrast, in the choice-not-changed cases these two regions showed the same correlation during the second decision. The preference choice did not change in subsequent decisions when the following two conditions were met: (i) the caudate exhibited a high activity at first sight and (ii) the OFC and the hippocampus showed a high functional correlation during the first preference decision. However, if one of these two conditions was not met during the first decision, the subsequent preference decision tended to be different. In principle, we may be able to determine the likelihood of changing mind by examining the activation of the caudate at the time of the first sight, because the caudate appears to 'know' if one changes one's mind.

MATERIALS AND METHODS

Participants

Thirty-six healthy volunteers gave written informed consent for this study, approved by the ethics committee of Tamagawa University. Participants were divided into three groups: Main Manip. group [Face Preference group with effective Gaze Manipulation; eight female, four male, age = 19.2 ± 0.38 (mean ± standard error of mean, SEM)], No Gaze Shift group (Face Preference group with ineffective Gaze Manipulation; six female, six male, age = 20.0 ± 0.26), and Roundness group (Face Roundness group with effective Gaze Manipulation; six female, six male, age = 19.9 ± 0.33). The No Gaze Shift and Roundness groups were used as the control groups to examine the effects of gaze manipulation.

Stimuli

One-hundred and sixty different faces (80 female, 80 male) were generated with a computer program (FaceGen; Singular Inversions). All images were presented on a 19-inch screen at 1024 × 768 pixel resolution. The viewing distance was always 57 cm. Eye movements were tracked with an eye tracking system (Arrington Research).

Prerating of faces

Before scanning, the Main Manip. and the No Gaze Shift groups were first asked to rate the facial attractiveness of the 80 female and 80 male faces on a scale from 1 to 7; the Roundness group rated the facial roundness of each face in the same way. After obtaining the ratings, faces were paired according to each participant's rating so that faces with close ratings were paired.

Instructions

Participants in the Main Manip. and No Gaze Shift groups were asked 'Who would you like to approach and talk to?' and participants in Roundness group were asked 'Which face is rounder than the other?'.

Task design and gaze manipulation

All participants underwent two functional magnetic resonance imaging (fMRI) scans, each consisting of 40 trials. A trial consisted of two-alternative forced-choice sessions (initial and final choice phases) and a gaze manipulation session (Figure 1, Supplementary Figures S1A and S1B). After the initial choice phase in each trial, participants performed the gaze manipulation and then the final choice phase. To randomize the task, the manipulation and the final choice phase were skipped in ~40% of all trials. In the choice phases, two faces appeared sequentially on the screen, and the participants were asked to choose the face according to the instructions by pressing a button within five stimulus presentation cycles. The cycle in which the participants pressed the button was termed the 'response cycle', and the immediately preceding cycle was termed the 'opening cycle'. The duration times of the fixation, the faces, and 'Which?' were 1500–3000, 50 and 1000 ms, respectively. In the gaze manipulation session, the Main Manip. and Roundness groups were exposed to the effective gaze manipulation and the No Gaze Shift group to the ineffective gaze manipulation. In the effective gaze manipulation, the two faces appearing in the initial choice phase were displayed six times each on the right or left side of the screen sequentially in random order, and the presentation time of each face was determined by the participant's choice in the initial choice phase. The presentation time of the chosen face in the initial choice session was 300 ms, and that of the unchosen face was 900 ms. In contrast, in the ineffective manipulation, the two faces were displayed six times each on the center of the screen sequentially in alternate order, and the presentation time of the two faces was 600 ms.

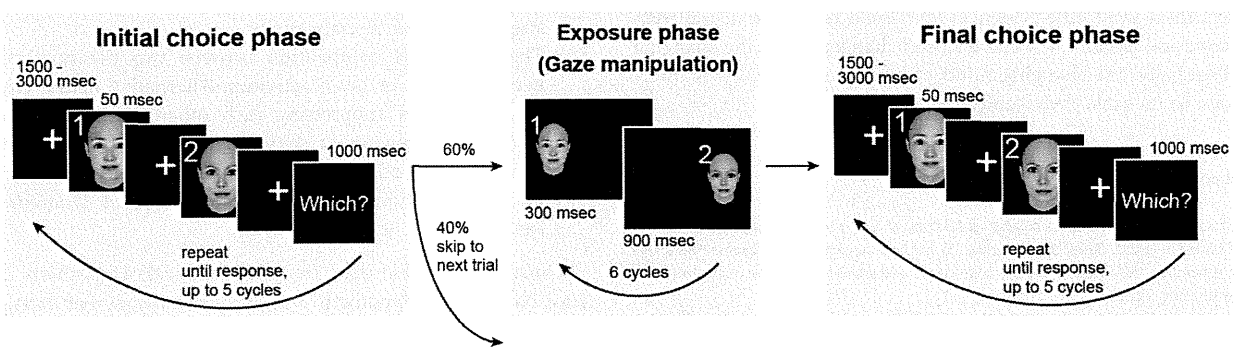

Fig. 1 Iterated choice task design. A trial, using the same face pair throughout, consisted of an initial choice phase, a manipulated exposure phase, and a final choice phase. To randomize the task, 40% of trials ended after the initial choice. In choice phases, the two faces were presented in sequence, followed by an option to respond. The cycle was repeated, up to five times, until the subject indicated which person they would rather 'approach and talk to'. In the exposure phase, durations were biased toward the initially unchosen face to subconsciously bias subject preferences.

Changing ratio

After the scans, we counted the number of the changing choices between the first choice session and second choice session, and calculated the changing ratio of the choice in each subject.

Imaging procedures

Functional imaging was conducted on a 3-Tesla Siemens Trio Tim MRI scanner. For each participant, we acquired whole-brain T1-weighted anatomical scans and gradient echo T2 weighted echo planar images (EPI) with BOLD contrast (TR = 2000 ms; TE = 25 ms; slice gap, 0.6 mm; FOV, 192 mm; slice thickness, 3.0 mm; 34 oblique axial slices). We used a tilted acquisition sequence at 30° to the AC–PC line to recover signal loss in the medial orbitofrontal cortex (mOFC; Deichmann *et al.*, 2003). The first 5 volumes of images were discarded to allow for equilibration effects.

Imaging data analysis

Image data were analyzed by using SPM8 (Wellcome Department of Imaging Neuroscience, Institute of Neurology, London, UK). To correct for participants' motion, the images were realigned to the mean volume image and spatially normalized to a bias-corrected T1 image, and spatial smoothing was applied by using a Gaussian kernel with a full width at half maximum (FWHM) of 8 mm.

Trials were separated into Firm Choice and Changed Choice trials on the basis of whether the initial and final choices matched or differed. We sorted the trials by the number of cycles required to make decisions and selected only the trials with two cycles to examine the temporal fluctuation of the neural activities and to compare the data with a previous report (Kim *et al.*, 2007). All MRI data in this article were limited to the Main Manip. group's data because the number of the Changed Choice trials in the other two groups were very small; thus, we could not detect any statistical differences in neural activity among the two groups. Each face presentation was treated as an event and categorized into chosen and unchosen faces as well as whether it was part of a Firm or Changed Choice trial. To avoid confusion, we refer to the faces on the basis of their status having been chosen or unchosen in the *final* choice phase.

Linear contrasts of regression coefficients (parameter estimates) were computed at the individual participant level in contrast to the final choice or the other face. The results from each participant were taken to a random effects level by including contrast images from each single participant into a paired *t* test. A statistical threshold at $P < 0.001$ or < 0.005 (uncorrected) was used.

Region of interest extraction

We used the MarsBar tool for SPM (http://marsbar.sourceforge.net/) to extract activations from the spherical regions of interest (ROIs) centered on the peak coordinates for the significant caudate, hippocampus and OFC contrasts (arrows in Figures 3 and 4).

Correlation analysis

In phases, where significant activity was found in the hippocampus and OFC, we used SPSS (IBM) software to run Pearson correlation analysis between contrast levels in the ROIs of the OFC and hippocampus.

Statistical analysis

All statistical analyses were performed with SPSS software.

RESULTS

Behavioral results

One procedural complication posed by compressed laboratory experiments such as this one is the typical paucity of reported changes due to consistency bias and rote behavior patterns. Thus, during the middle exposure phase, we employed a gaze manipulation paradigm, which has been shown to subliminally bias subject preferences (Shimojo *et al.*, 2003). Control experiments with unmanipulated exposures and a 'roundness' choice task verified that this manipulation was successful in leading subjects to change their minds; however, only in the preference task ($P < 0.001$, Supplementary Figure S1C).

The temporal dynamics of the changing mind

We analyzed fMRI activity in response to face presentations by performing comparisons for the face of *final* choice *vs* the other face. These comparisons were performed for data from the first stimulus presentation cycle (opening cycle) and the last cycle prior to choice response (response cycle) separately (Figures 2, 3, 4, Supplementary Figures S2 and S3, and the main effects were shown in Supplementary Tables S1–S4). We divided the trials into Firm Choice trials, where subjects made the same choice in both phases, and Changed Choice trials, where the initial and final choices differed. As shown in Supplementary Figures S2 and S3, these final chosen *vs* unchosen face comparisons were statistically significant regardless of the size choice of ROIs.

The effects of gaze manipulation on the face preference decision

To examine the effects of gaze manipulation on the preference decision, we compared the contrast between the initial choice and the final choice phase. The fMRI data revealed that there was an increase and inversion of the activity in the hippocampus and OFC only in the Changed Choice trials but not in the Firm Choice trials (Figure 2).

The temporal dynamics of caudate activity during changing of mind

The caudate showed a significant contrast toward the face of the final choice in the opening cycles of both choice phases, and this was true for both Firm and Changed Choice trials (Figures 3A and 4A). The caudate activity exhibited a dynamic spatial pattern as the trial progressed. Significant activity began in the middle caudate and then moved to the anterior caudate (Figures 3A, D, 4A, B, D, and Supplementary Figure S4). In the Firm Choice trials, this pattern was more distinct with the middle caudate being significant only at the opening of the initial choice phase and the anterior caudate being significant only at the opening of the final choice phase. The Changed Choice trials showed a more sustained version of the same pattern (Figure 4B and D). The middle caudate extended its significant activity throughout the initial choice phase and into the opening of the final choice phase. The pattern of the anterior caudate began to increase in the initial choice phase and became significant at the opening of the final choice phase (Figures 3D and 4D).

The functional correlation of the hippocampus and the OFC

The hippocampus and the OFC (Figures 3B, C, 4C, E and F) also showed significant contrast at selected phases and during the 'Which' response, which is the time to decide which face is preferable to the other and indicate it by pressing a button. We calculated the correlations between these regions during each choice phase. We found significant activity and correlations in phases where subjects made a firm or final choice, but not in phases where their choice was tentative

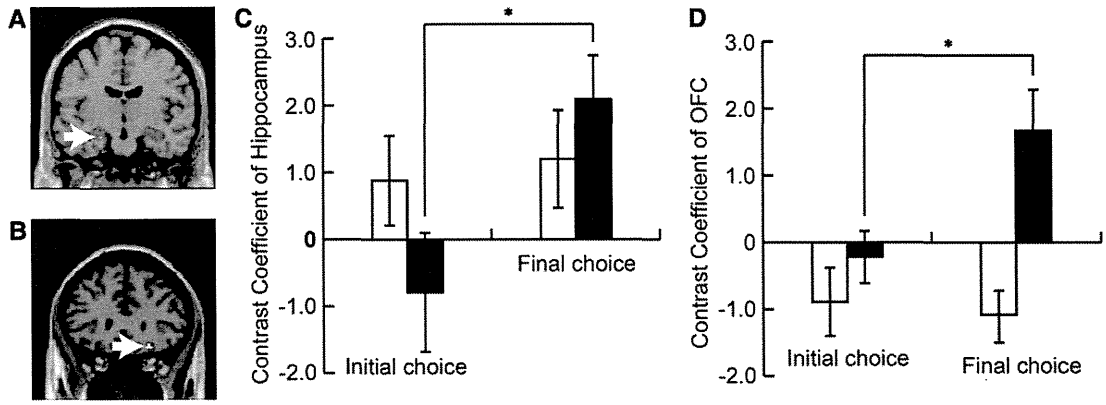

Fig. 2 Gaze manipulation increased the activation of the hippocampus and the OFC. The contrast between the response cycles of the initial and final choice phase showed activations in the left hippocampus (**A**, −22, −12, −20) and the right OFC (**B**, 24, 30, −16). ROI analysis showed that these regions were activated only in the Changed Choice trials (**C**, **D**). These two regions were considered to play an important role in face attractiveness judgment. White arrows indicate the peak voxels in the hippocampus (**A**) and the OFC (**B**). White bars are Firm Choice, and black bars are Changed Choice trials. To define the activated regions, a statistical threshold at $P < 0.005$ (uncorrected) was used. Each peak voxel was used as the center of 4-mm radius spherical ROI. Error bars mean SEM; $^{*}P < 0.05$. All MRI data are from the Main Manip. group.

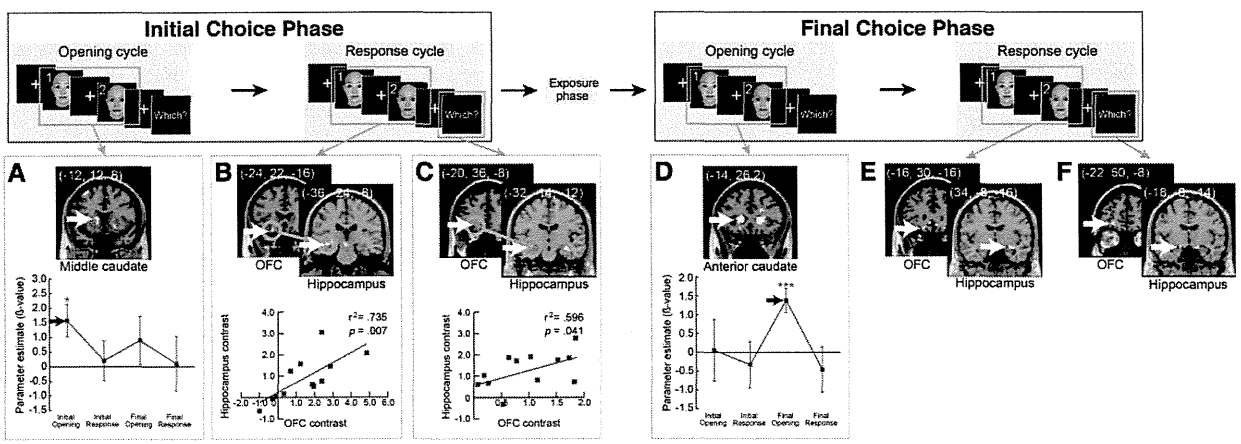

Fig. 3 Significant activity contrasts and correlations at successive cycles of the Firm Choice trials. Initial choice phase and Final choice phase trials were analyzed separately. White arrows indicate the peak voxels in the caudate, the hippocampus, and the OFC. (**A**, **D**) The bottom graphs show the time course of caudate activity. (**B**, **C**) Blue circles indicate a significant correlation of contrast levels between the hippocampus and OFC, and subject-by-subject plot of contrast levels in hippocampal *vs* OFC ROIs, indicating significant correlations. To define the activated regions, a statistical threshold at $P < 0.005$ (uncorrected) was used. $^{***}P < 0.001$, $^{**}P < 0.005$, $^{*}P < 0.01$.

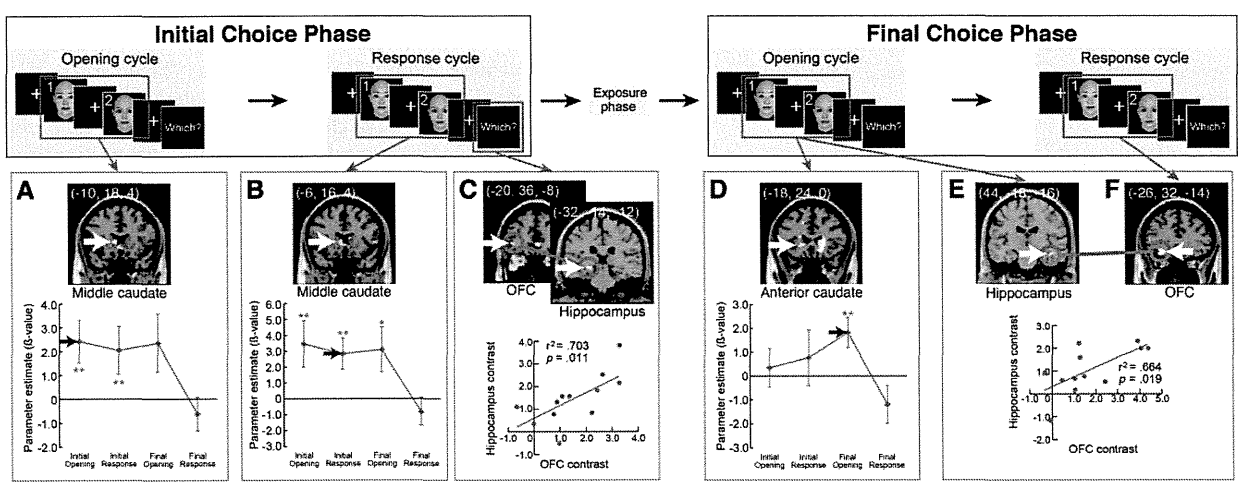

Fig. 4 Significant activity contrasts and correlations at successive cycles of the Changed Choice trails. Initial choice phase and Final choice phase trials were analyzed separately. White arrows indicate the peak voxels in the caudate, the hippocampus and the OFC. (**A**, **D**) The bottom graphs show the time course of caudate activity. (**C**, **E**, **F**) Red circles indicate a significant correlation of contrast levels between the hippocampus and OFC, and subject-by-subject plot of contrast levels in hippocampal *vs* OFC ROIs. To define the activated regions, a statistical threshold at $P < 0.005$ (uncorrected) was used. $^{***}P < 0.001$, $^{**}P < 0.005$, $^{*}P < 0.01$.

and was going to change. In the Firm Choice trials, significant contrast and correlation was present in the initial choice phase (Figure 3B and C, blue circles and bottom scatter-plot). In the final choice phase, where subjects stuck to their earlier choice, there was also a significant contrast in both areas (Figure 3E and F), but they lacked the correlation between the hippocampus and the OFC. In Changed Choice trials, there was no significant contrast in the response cycle in the initial choice phase in which the subjects responded with a choice that was later reversed. However, there was a correlation between the hippocampus and the OFC in the 'Which' response timing (Figure 4C). Significant contrast and correlation was present in the final choice phase (Figure 4E and F, red circles and bottom scatter-plot). Finally, we also found a significant contrast in the ACC and the DLPFC; however, only in the Changed Choice trials (Supplementary Figure S5).

DISCUSSION

We identified the neural basis underlying the changing of mind, and the dynamic process appears be as follows: At first sight (initial opening cycle), (i) the caudate is activated in response to 1 face (we call it 'Caudate-dominant face') over the other ('Caudate-nondominant face'). In other words, caudate activity reflects the final choice even at the very beginning of the trial, even when the subject is about to make an initial choice favoring the opposite face. This might lead to (ii) a functional correlation between the hippocampus and the OFC, which provides the conditions for the choice. The Caudate-dominant face is typically selected for the initial choice; however, occasionally the Caudate-nondominant face is selected possibly due to noise or fluctuations in competition. Indeed, only in the Choice Changed trials, the ACC and the DLPFC, which are well-known regions to correlate with a decision conflict and resolution, (MacDonald *et al.*, 2000; Milham *et al.*, 2001; Pochon *et al.*, 2008) showed high activation (Supplementary Figure S5). While we admit that this is a *post hoc* interpretation of the activity, the known functions of these areas and the specificity to the Choice Changed trials seem to be highly consistent.

It is likely that when (i) and (ii) above are satisfied, the choice is maintained in the final judgment. When these conditions are not met, a change of mind is likely to occur. Interestingly, even in such cases, there is an increase in the activity of both the hippocampus and the OFC as well as in the functional correlation in response to a newly chosen face. However, there were some laterality differences in the connected regions (Figures 3B, C, 4C, E and F) and the increase in activity (Figure 2). This is logical if we consider that the second choice is firmer than the first one in the Choice Changed trials. Either way, our findings add a significant constraint on the manipulation effect—it reverses the choice, only (or mostly) when the neural response to the finally chosen object is high in the memory-related brain circuits (such as the hippocampus–OFC network) from the beginning. Meanwhile, the transitions from the middle to the anterior caudate (Figures 3A, D, 4A, B and D) might reflect a progression in consolidation of a preference tendency into firm preferences. Along with the shift in caudate activity, the hippocampal and OFC activity might also possibly be involved in the consolidation of a preference tendency into a firm decision. Alternatively, the activity could reflect ancillary processes such as a memory process that was engaged because of the decision (Tsukiura and Cabeza, 2011).

More precisely, our findings provide a new insight into the temporal neural processing during face preference decisions, especially regarding the function of the caudate, OFC and hippocampus. The ventral striatum and the OFC are two major subcortical and cortical regions, respectively, which have been strongly implicated in reward-related processing (Knutson *et al.*, 2001; Cardinal *et al.*, 2002; O'Doherty *et al.*, 2001). In general, the ventral striatum is involved in encoding errors in predicting future rewards (i.e. the reward anticipation), whereas the OFC is involved in encoding stimulus-reward value and in representing expected future rewards (O'Doherty *et al.*, 2004). The present data indicate a distinct contribution of these two reward-related regions in terms of temporal dissociation and consolidation during face preference decision-making. As shown by the caudate activity at first sight (the opening cycle of the initial choice phase), the relative evaluation of two faces was made instantly without any delay. The information was then transferred from the caudate to the OFC and the hippocampus, which are anatomically connected (Barbas and Blatt, 1995; Carmichael and Price, 1995; Lavenex *et al.*, 2002). This functional connectivity between the OFC and hippocampus during face preference judgment reflects the positive signals generated by an attractive face (Tsukiura and Cabeza, 2011) and is possibly involved in the consolidation of a preference tendency into a firm decision. The current study demonstrates such a serial transfer of the preference signal from the subcortical level to the cortical level for the first time. At the same time, it also indicates that such a signal transfer often does not work; for example, in the Choice Changed cases. In a sense, the process of changing of mind is very simple; signal transfer is disrupted by some conflicting signals representing the activity of the ACC and the DLPFC, and the participant's choice is not consistent with a Caudate-dominant face. In such cases, the choice is changed rapidly.

Moreover, our data may be interpreted as, changing of mind does not really occur in the narrower sense. It can be argued that the future decision has already been made implicitly at first sight and never been changed (as indicated by the early caudate activity). It may be because of random noise or fluctuation in the downstream processes, or overridden by the cortical deliberation system and, therefore, the 'dominant' face was not chosen in the initial decision. In the final decision, the original evaluation is transferred from the caudate (implicit, subcortical level) to the hippocampus and OFC (explicit, cortical level) because of the gaze manipulation, and this transfer enables the so-called changing of the mind. This interpretation is different from a more integrated/consistent view of decision making, but rather reminiscent of the 'neural competition of choices' idea (Reynolds *et al.*, 1999; Resulaj *et al.*, 2009). The results may be applicable to various other cases of decision changes in the laboratory and in everyday life.

SUPPLEMENTARY DATA

Supplementary data are available at SCAN online.

Conflict of Interest

None declared.

REFERENCES

Albantakis, L., Deco, G. (2011). Changes of mind in an attractor network of decision-making. *PLOS Computational Biology, 7*, e1002086.

Barbas, H., Blatt, G.J. (1995). Topographically specific hippocampal projections target functionally distinct prefrontal areas in the rhesus monkey. *Hippocampus, 5*, 511–33.

Cardinal, R.N., Parkinson, J.A., Hall, J., Everitt, B.J. (2002). Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. *Neuroscience and Biobehavioral Reviews, 26*, 321–52.

Carmichael, S.T., Price, J.L. (1995). Limbic connections of the orbital and medial prefrontal cortex in macaque monkeys. *The Journal of Comparative Neurology, 363*, 615–41.

Deichmann, R., Gottfried, J.A., Hutton, C., Turner, R. (2003). Optimized EPI for fMRI studies of the orbitofrontal cortex. *Neuroimage, 19*, 430–41.

Kable, J.W., Glimcher, P.W. (2009). The neurobiology of decision: consensus and controversy. *Neuron, 63*, 733–45.

Kim, H., Adolphs, R., O'Doherty, J.P., Shimojo, S. (2007). Temporal isolation of neural processes underlying face preference decisions. *Proceedings of the National Academy of Sciences, 104*, 18253–8.

Knutson, B., Adams, C.M., Fong, G.W., Hommer, D. (2001). Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *Journal of Neuroscience, 21*, RC159.

Krajbich, I., Armel, C., Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience, 13*, 1292–8.

Lavenex, P., Suzuki, W.A., Amaral, D.G. (2002). Perirhinal and parahippocampal cortices of the macaque monkey: projections to the neocortex. *The Journal of Comparative Neurology, 447*, 394–420.

MacDonald, A.W., 3rd, Cohen, J.D., Stenger, V.A., Carter, C.S. (2000). Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science, 288*, 1835–8.

Mende-Siedlecki, P., Cai, Y., Todorov, A. (2013). The neural dynamics of updating person impressions. *Social Cognitive and Affective Neuroscience, 8*, 623–31.

Milham, M.P., Banich, M.T., Webb, A., et al. (2001). The relative involvement of anterior cingulate and prefrontal cortex in attentional control depends on nature of conflict. *Cognitive Brain Research, 12*, 467–73.

O'Doherty, J., Kringelbach, M.L., Rolls, E.T., Hornak, J., Andrews, C. (2001). Abstract reward and punishment representations in the human orbitofrontal cortex. *Nature Neuroscience, 4*, 95–102.

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., Dolan, R.J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science, 304*, 452–4.

Pennartz, C.M., Berke, J.D., Graybiel, A.M., et al. (2009). Corticostriatal interactions during learning, memory processing, and decision making. *Journal of Neuroscience, 29*, 12831–8.

Pochon, J.B., Riis, J., Sanfey, A.G., Nystrom, L.E., Cohen, J.D. (2008). Functional imaging of decision conflict. *Journal of Neuroscience, 28*, 3468–73.

Reynolds, J.H., Desimone, R. (1999). The role of neural mechanisms of attention in solving the binding problem. *Neuron, 24*, 19–29.

Resulaj, A., Kiani, R., Wolpert, D.M., Shadlen, M.N. (2009). Changes of mind in decision-making. *Nature, 461*, 263–66.

Shimojo, S., Simion, C., Shimojo, E., Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nature Neuroscience, 6*, 1317–22.

Tsukiura, T., Cabeza, R. (2011). Remembering beauty: roles of orbitofrontal and hippocampal regions in successful memory encoding of attractive faces. *Neuroimage, 54*, 653–60.

OPEN ACCESS Freely available online

PLOS | ONE

How Sound Symbolism Is Processed in the Brain: A Study on Japanese Mimetic Words

Junko Kanero[1☯], Mutsumi Imai[2☯], Jiro Okuda[3], Hiroyuki Okada[4], Tetsuya Matsuda[5]*

1 Department of Psychology, Temple University, Philadelphia, Pennsylvania, United States of America, **2** Faculty of Environment and Information Studies, Keio University at Shonan-Fujisawa, Endo, Fujisawa, Kanagawa, Japan, **3** Department of Intelligent Systems, Faculty of Computer Science and Engineering, Kyoto Sangyo University, Kamigamo-Motoyama, Kita-Ku, Kyoto, Japan, **4** Department of Mechanical Systems, College of Engineering, Tamagawa University, Machida, Tokyo, Japan, **5** Tamagawa University Brain Science Institute, Machida, Tokyo, Japan

Abstract

Sound symbolism is the systematic and non-arbitrary link between word and meaning. Although a number of behavioral studies demonstrate that both children and adults are universally sensitive to sound symbolism in mimetic words, the neural mechanisms underlying this phenomenon have not yet been extensively investigated. The present study used functional magnetic resonance imaging to investigate how Japanese mimetic words are processed in the brain. In Experiment 1, we compared processing for motion mimetic words with that for non-sound symbolic motion verbs and adverbs. Mimetic words uniquely activated the right posterior superior temporal sulcus (STS). In Experiment 2, we further examined the generalizability of the findings from Experiment 1 by testing another domain: shape mimetics. Our results show that the right posterior STS was active when subjects processed both motion and shape mimetic words, thus suggesting that this area may be the primary structure for processing sound symbolism. Increased activity in the right posterior STS may also reflect how sound symbolic words function as both linguistic and non-linguistic iconic symbols.

Citation: Kanero J, Imai M, Okuda J, Okada H, Matsuda T (2014) How Sound Symbolism Is Processed in the Brain: A Study on Japanese Mimetic Words. PLoS ONE 9(5): e97905. doi:10.1371/journal.pone.0097905

Editor: Hisao Nishijo, University of Toyama, Japan

Received February 26, 2014; **Accepted** April 25, 2014; **Published** May 19, 2014

Funding: This work was supported by MEXT (Ministry of Education, Culture, Sports, Science and Technology) Kakenhi grant to MI (#15300088) and to TM (#21300301), MEXT Grant-in-Aid for Scientific Research on Innovative Areas: Prediction and Decision Making to MI (#23120003), and MEXT Global Center of Excellence (GCOE) program awarded to Tamagawa University. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist. In Figure 1, the authors present pictures of people walking. This person is the first author of this manuscript.

* E-mail: tetsuya@lab.tamagawa.ac.jp

☯ These authors contributed equally to this work.

Introduction

Traditional linguistics assumes that language is independent from perceptual, motor, or affective experience and that pairings between a word's sound and its meaning are arbitrary [1]. The notion of sound symbolism, however, challenges this well-accepted belief by suggesting natural and systematic relationships between word sound and meaning [2]. People across the world intuitively associate the nonsense word "baluma" to a round shape and "takete" to a spiky shape (i.e., bouba/kiki effect) [3–4]. Since then, a large body of linguistic and psychological research has investigated sound symbolism (e.g., [5]). Sound symbolic words are found in many languages including English. For example, *bump* and *thump* have sounds similar to their meanings–an event with an abrupt end [6]. Furthermore, a number of languages, including Japanese, have a large grammatically defined word class in which sound symbolism is apparent. These sound symbolic words, which are called mimetics, idiophones, or expressives, are abundant in African [7] and East Asian languages [8–14]. Adults [15–16], as well as infants and toddlers [17–22], are sensitive to sound symbolism in mimetic words, regardless of the language they speak. For example, the sound symbolism of Japanese mimetic words promotes verb learning in both Japanese- and English-reared children [18–20]. The existence of sound symbolism across languages has led some researchers to claim that this phenomenon can provide insights into the ontogenesis and phylogenesis of language [4,18,23]. Despite its significance, the neural mechanisms of sound symbolism are yet to be sufficiently investigated.

Ramachandran and Hubbard [4] hypothesized that sound symbolism shares the neural mechanisms underlying synesthesia. They further argue that multi-sensory integration at the temporal–parietal–occipital (TPO) junction, or more specifically the angular gyrus, is the critical region for sensing sound symbolism. In addition, they noted anecdotally that individuals with damage to the angular gyrus did not show the bouba/kiki effect. Nevertheless, these ideas are largely speculative and have never been investigated empirically.

We agree with this previous hypothesis that perceiving sound symbolism requires a unique integrative process. We hypothesize, however, that the posterior part of the superior temporal sulcus (STS) is a key area in this processing. The STS represents 2 routes for conceptual access: the left STS processes linguistic sounds, whereas the right STS processes environmental sounds [24]. The universal understanding of mimetic words suggests that these words possess some features of non-linguistic environmental sounds that do not require language system for understanding. We argue that neural processing of sound symbolic words

integrates the two conceptual processes involving the bilateral STS.

A previous functional magnetic resonance imaging (fMRI) study found that auditory presentation of Japanese mimetic words for animal sounds (e.g., *ka-ka*, onomatopoeia for crow croaks) more strongly activated the right STS than the names of the animals (e.g., *karasu*, "crow" in English) [25]. Similarly, Japanese mimetic words for animal sounds more strongly activated the STS bilaterally than the actual animal sounds (e.g., sound of a crow croaking). That study concluded that onomatopoeic words activate both the left and right STS because they have acoustic properties similar to real animal sounds. The acoustic similarity between mimetic words and the actual sound, however, cannot fully explain the phenomenon of sound symbolism, because sound symbolic words are not limited to mere mimicry of environmental sounds.

For example, Japanese mimetic words are roughly classified into 3 categories–phonomimes, phenomimes, and psychomimes [26]. Phonomimes, or *giongo*, are onomatopoeia that acoustically imitate actual sound (e.g., *wanwan* for dog barking). Phenomimes, or *gitaigo*, represent the characteristics of input from non-auditory senses (e.g., *yotayota* for walking clumsily). Psychomimes, or *gijogo*, represent psychological states (e.g., *wakuwaku* for the feeling of excitement). Several studies demonstrated that Japanese as well as non-Japanese speakers can discern sound-meaning correspondences in the latter two types of mimetics [18–20,27,28]. Sound symbolism in English, such as *squeeze*, *squirt*, *squint*, *bump*, *thump*, and *plump* [6], are found beyond the non-auditory domain as well. Thus, in order to fully understand the neural processing of sound symbolism, we must investigate sound symbolism in the non-auditory domain.

We hypothesize that right STS participation can differentiate sound symbolic words from non-sound symbolic words. Therefore, all types of mimetic words, including phenomimes and psychomimes, should activate the right STS. To determine whether the right STS is the primary structure for sound symbolism processing, we investigated whether this region responds to non-onomatopoeic mimetic words. For this purpose, we tested mimetics in two domains, motion and shape, and all words were presented visually rather than auditorily. Experiment 1 contrasted Japanese mimetic words with non-sound symbolic conventional verbs and adverbs, all of which express aspects of human motion. Experiment 2 compared the neural processing of mimetic words for human motion as well as for shape to ensure that the right STS activation is not limited to the domain of motion. Interpretation of the STS activation in Experiment 1 requires caution because the STS shows activation during the processing of animated figures [29,30] and point-light biological motion [31]. If the right STS is the key structure for sound symbolism processing, we should see the activation of this area both for motion mimetic words and for shape mimetic words. Experiment 2 tested mimetic words only, as differences in brain activation across word classes (mimetic words, verbs, and adverbs) were demonstrated in Experiment 1, and as with inclusion of multiple word classes would substantially increase the length of each scanning session.

Materials and Methods

Experiment 1

Participants. Sixteen native Japanese speakers aged 22–25 years (7 women, 9 men; mean age = 23.7 years) participated in this study. All participants were right-handed, had normal or corrected-to-normal vision, and had no history of neurological or psychiatric symptoms. Data from 5 participants were excluded due to artifact (e.g., head movements >3 mm) or inadequate task performance (e.g., failing to press buttons as instructed during scanning sessions); thus data were analyzed from the remaining 11 participants (4 women, 7 men; age range = 22–25 years; mean age = 23.4 years). The individual in this manuscript has given written informed consent to publish these case details. The study was approved by the ethics committee of Tamagawa University.

Design and procedure. Stimuli were 16 video clips of a human agent moving from left to right in different manners. Each video clip was 5-sec long, and was presented simultaneously with a sound symbolic mimetic word, a non-sound symbolic adverb, or a non-sound symbolic verb. All words were presented at the bottom of the video in *hiragana* (a type of Japanese orthographical coding in which each character represents a syllable). In half of the trials, the word and manner of motion semantically matched, whereas in the other half, the items were mismatched (e.g., the verb *aruiteiru* "to walk in the progressive aspect," was shown with a video clip of an agent skipping). Thus, for each word class (mimetic words, verbs, or adverbs), 8 motion-word pairs were matched and 8 pairs were mismatched. Participants were instructed to determine the degree of match between the word and the motion as the video clips were presented. After each video clip, a fixation point appeared on the screen for 3 sec, and participants indicated the degree of match between the word and the motion on a scale of 1 to 5 by pressing the appropriate button with a right-hand finger (Figure 1). As Experiment 1 used a 1.5 scanner, we used a block design to maximize sensitivity to the brain response: 4 blocks were presented for each word class (mimetic words, verbs, or adverbs), with each block consisting of 4 motion-word pairs from the same word class. The order of the blocks was rotated among participants. A fixation point was inserted for 10 sec at the end of each block.

Stimuli and stimulus validation. Three pretests examined 120 preselected words to ensure that the mimetic words, verbs, and adverbs were balanced in terms of imageability, familiarity, and age of acquisition (AOA). Twenty-eight participants who were native-Japanese speakers rated how imageable each word was on a scale of 1 to 7. Twenty-seven participants categorized word familiarity on a scale from 1 to 7. Twenty-two participants were asked to indicate the approximate age at which they learned words from the following 8 categories: infancy, preschool, first to third grade, fourth to sixth grade, junior high school, high school, university or college, or do not know the meaning. The pretest results indicated significant differences among the 3 word classes with respect to imageability (mimetic words: 5.28; verbs: 6.40; adverbs 5.62; $F(2,81) = 3.11$, $p<0.05$) and familiarity (mimetic words: 5.42; verbs: 6.51; adverbs: 6.08; $F(2,78) = 3.11$, $p<0.05$); although mimetic words and adverbs did not significantly differ in imageability ($t(27) = 1.200$, $p = 0.241$). The results of the AOA survey indicated that participants acquired mimetic words and verbs earlier than adverbs (mean rating scores were 1.52 for mimetics, 1.55 for verbs, and 2.93 for adverbs); however, no significant difference was found between AOA of the mimetic words and verbs (Freedman test, $p = 0.76$).

Materials and imaging parameters. Imaging was performed using a 1.5-T MRI scanner (SIEMENS MAGNETOM SONATA, Erlangen, Germany). A high-resolution (1×1×1 mm) T1-weighted anatomical reference image was acquired from each participant using a rapid acquisition gradient echo (MP-RAGE) sequence. Multi-slice gradient echo planar imaging (EPI) was used with a TE of 50 ms and a TR of 2000 ms. Slice-acquisition was ascending within the TR interval. The matrix acquired was 64×64 voxels with a field of view of 192 mm, resulting in an in-plane resolution of 3 mm. Slice thickness was 3 mm (20 slices, whole brain coverage).

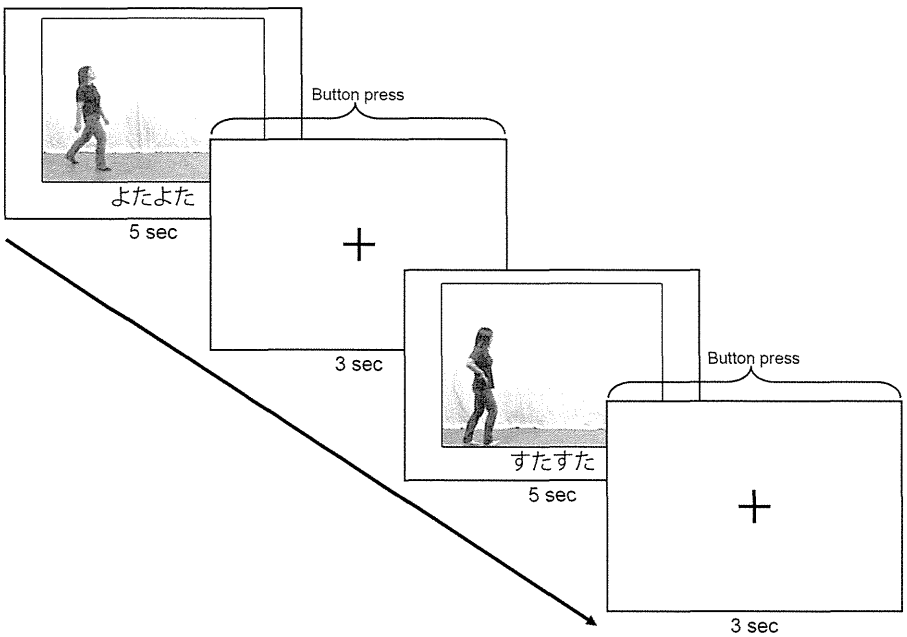

Figure 1. Study paradigm for Experiment 1. Experiment 1 used a blocked design. A 5-sec video clip presented a person moving from left to right and a matched/mismatched word that were followed by a 3-sec presentation of a fixation point. Participants were asked to press a button during the fixation point presentation to indicate the degree of match, on a scale of 1–5, between the motion and the mimetic words. This example shows two trials in the mimetic word block. The mimetic words depicted in this example: よたよた (*yotayota*) "walk clumsily" and すたすた (*sutasuta*) "walk very quickly".
doi:10.1371/journal.pone.0097905.g001

fMRI data analyses. fMRI data were analyzed using SPM8 software (Wellcome Department of Imaging Neuroscience, Institute of Neurology, London, UK). The gradient-echo echo-planar images for each time series were realigned with reference to the first image acquired in each session to correct for head motion. The anatomical images were co-registered with the mean functional images and normalized to the Montreal Neurological Institute (MNI) brain template. Functional data were normalized using the same transformation parameters and smoothed in the spatial domain (isotropic Gaussian kernel of 8 mm full width half-maximum). Low-frequency drifts were removed using a high-pass filter [32], and a first order autoregressive model (AR1) [33] was applied for eliminating the temporal autocorrelation of the fMRI time series data.

The fMRI time series for each participant were analyzed using a block design approach with a general linear model. The images were sorted by trial type (matched and mismatched trials), and regions unique to mimetic processing were calculated by subtracting verbs and adverbs from mimetic words. The vectors indicating the onset and duration of each of the 3 word classes (mimetic words, verbs, and adverbs) were convolved with a hemodynamic response function. The results for the single subject analyses were then used for group analyses. Images representing the estimated cerebral effects from the [mimetic words – verbs – adverbs] for each subject were analyzed using a one-way ANOVA to determine the consistency of the effects across subjects. To ensure that the activation patterns of mismatched motion-word pairs were different, the same procedure was conducted for mismatched pairs.

Experiment 2

Participants. Fifteen native Japanese speakers aged 17–26 years (8 women, 7 men; mean age: 20.93 years) participated in the fMRI study. All participants were recruited on the basis of the same criteria as in Experiment 1. Four participants were excluded from the analysis as not enough data were collected for these subjects (less than 10 trials in one condition). The final data set consisted of 5 women and 6 men (mean age: 21.13 years; range = 17–27 years).

Stimuli. One hundred and fourteen animation clips and their corresponding mimetic words were used in the main fMRI experiment. Each video clip depicted a simple line-drawing figure with hands and legs, and this "agent" either stayed still in the center of the screen or moved from left to right on a white background (Figure 2). The still and moving images were used for the shape and motion trials, respectively.

Twenty-four mimetic words referring to human motion and 35 mimetic words referring to shape were selected from a dictionary of Japanese mimetic words (*Giongo • Gitaigo 4500 Nihongo Onomatope Jiten*) [34]. Two separate rating tests, a web-rating test and a behavioral rating test, were conducted prior to the fMRI scanning to assure that the set of experimental stimuli contained both matched motion/shape-word pairs and mismatched motion/shape-word pairs. All participants who took part in the rating tests were native Japanese speakers who did not participate in Experiments 1 or 2.

In the web-rating test, 108 participants rated the degree of match between mimetic words and shapes/manners of motion on a scale of 1 to 5. 57 participants rated the degree of match between mimetic words and manners of motion, whereas 51 participants rated the degree of match between mimetic words and shapes. Each participant was presented with 105 pairs of words and their referents. From this analysis, 50 manners of motion and 48 shape figures were selected.

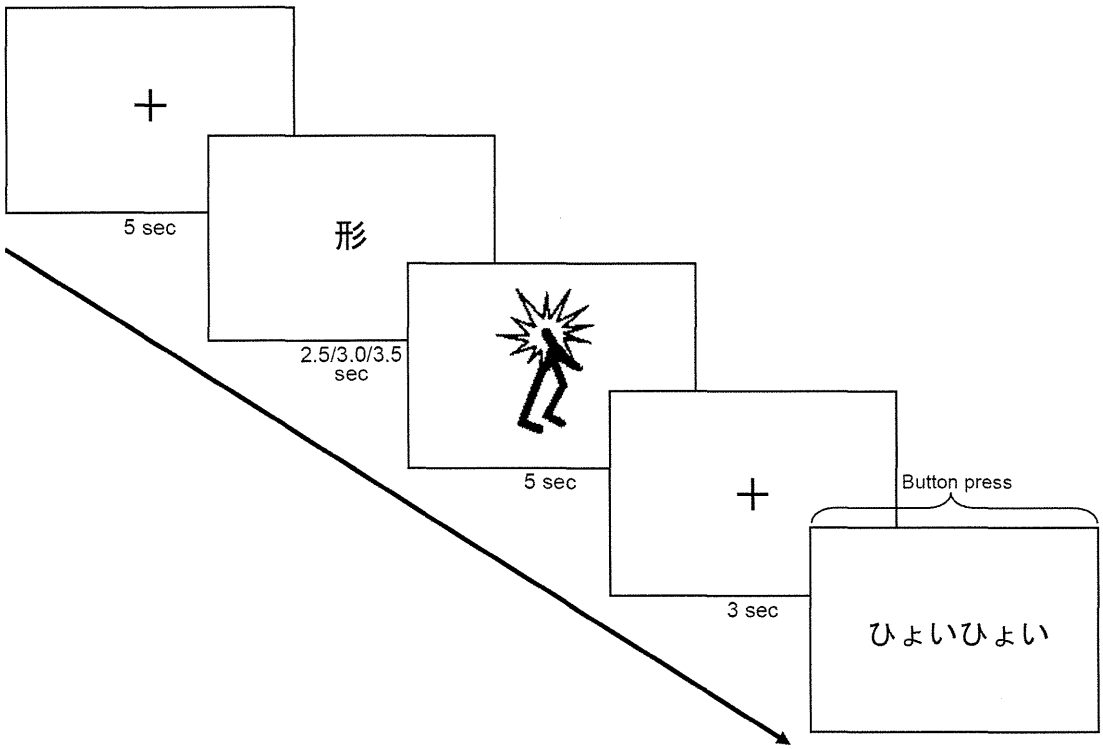

Figure 2. Study paradigm for Experiment 2. Experiment 2 used an event-related design. Stimuli were presented in the following order: 5-sec fixation point, 1-word instruction (presented either for 2.5, 3.0, or 3.5 sec) indicating the trial type (形 "shape" or 動き "motion"), 5-sec video clip, 3-sec fixation point, and a mimetic word. Video clips depicted an agent who stayed still in the shape trials and moved from left to right in the motion trials. During the presentation of a mimetic word, participants pressed a button to indicate the degree of match between the referent and the mimetic word. The mimetic word depicted in this example is ひよいひよい (*hyoihyoi*) which means "jumping effortlessly" in this context.
doi:10.1371/journal.pone.0097905.g002

After the web rating test, the remaining manners of motion and shapes were combined to create animation clips of an agent that was either motionless or that moved across the display, as described above.

In the behavioral rating test, 29 participants rated each stimuli pair (motion/shape and word) in the same manner as the scanning experiment. Thirteen shape/motion-word pairs that were judged as neither matched nor mismatched were excluded at this point. The final set of stimuli consisted of highly matched (mean rating score: 4.16 and 4.30 for the motion trials and shape trials respectively) and mismatched pairs (mean rating score: 1.34 and 1.29 for the motion trials and shape trials respectively). A total of 114 video clips (57 for each modality) were used in the fMRI experiment.

Design and procedure. Each shape or manner of motion appeared 1–8 times, and each shape-motion combination was different. Thus, participants saw each video clip once. A fixation point was presented for 5 sec, which was followed by a one-word instruction (either "motion" or "shape") that directed participants to attend to either the motion or shape of the agent in the animation clip. The duration of the instruction was jittered and was 2.5, 3, or 3.5 sec; the duration for all video clips was 5 sec. After each video clip, a sound symbolic mimetic word was visually presented. In some trials, the mimetic word and indicated visual property (motion or shape) were semantically matched, but these were mismatched in other trials (e.g., a hopping motion followed by the word *yotayota* "to walk clumsily"). Participants judged the degree of match between the manner of motion in motion trials and the shape of the agent and mimetic word in shape trials. Participants pressed 1 of 5 buttons while the mimetic word was on screen. 11 Stimuli sequences were presented in pseudo-random order to control the order effects, and all words were shown in *hiragana*.

Imaging parameters and analysis. Scanning was performed with a 3.0-T MRI scanner (Siemens MAGNETOM Torio-Tim, Erlangen, Germany). A high-resolution (1×1×1 mm) T1-weighted anatomical reference image was acquired from each participant using a rapid acquisition gradient echo (MP-RAGE) sequence. Multi-slice gradient echo planar imaging (EPI) was used with a TE of 25 ms and a TR of 2500 ms. Slice-acquisition was ascending within the TR interval. The matrix acquired was 64×64 voxels with a field of view of 192 mm, resulting in an in-plane resolution of 3 mm. Slice thickness was 3 mm (42 slices, whole brain coverage). The acquisition window was tilted at an angle of 30° relative to the AC-PC line in order to minimize susceptibility artifacts in the orbitofrontal cortex. The fMRI data were analyzed using SPM8 software and preprocessed using the steps described for Experiment 1.

We classified the trials as matched trials with high rating scores (4 or 5) or mismatched trials with low rating scores (1 or 2). Statistical analysis of the behavioral data was performed using 2 factors: Modality (motion/shape) and Degree of Match (matched/mismatched). Thus, the trials were divided into 4 cell means: Shape-High (shape trials with a high rating score), Shape-Low (shape trials with a low rating score), Motion-High (motion trials with a high rating score), and Motion-Low (motion trials with a low rating score). For fMRI analysis, we focused on highly

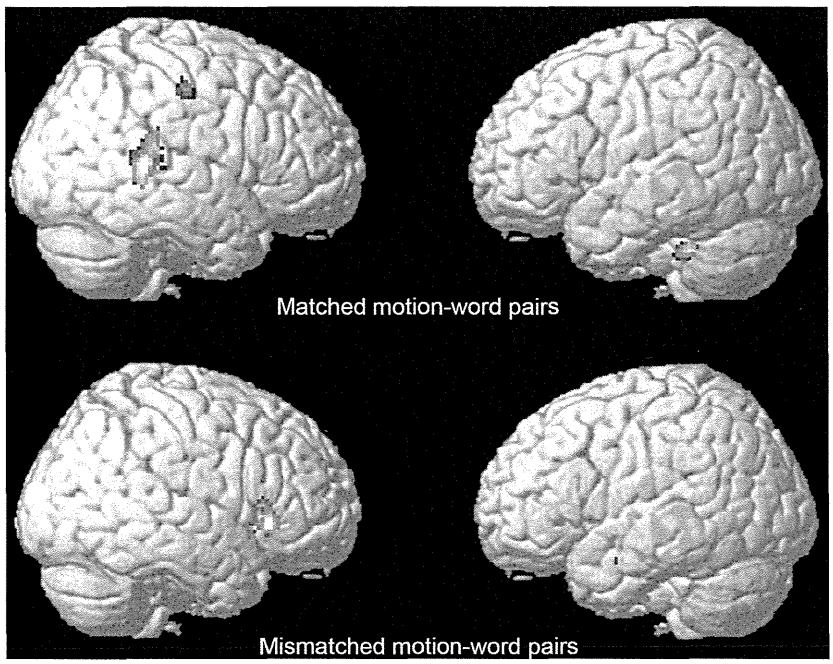

Figure 3. Results of Experiment 1. Regions that showed greater activation for mimetic words than for non-mimetic verbs and adverbs ($p<0.05$, FWE corrected; see Materials and Methods).
doi:10.1371/journal.pone.0097905.g003

matched word-referent pairs, and thus highly matched trials (Shape-High and Motion-High) and mismatched trials (Shape-Low and Motion-Low) were analyzed separately. General Linear Model EPI time series were analyzed using the general linear model function implemented in SPM8. At the first level (i.e., within subjects), Shape-High, Shape-Low, Motion-High, and Motion-Low were modeled separately, creating 4 regressors. At the second level (i.e., across subjects), a one-sample *t*-test was performed on each regressor to examine the activation level.

Results

Experiment 1

Behavioral results. We examined whether the rated degree of match between the word and motion itself were comparable across the 3 word classes. For each word class, the degree of match was high for the highly matched pairs (mimetic words: 4.24; adverb: 4.30; verb: 4.09) and low for the mismatched pairs (mimetic words: 1.65; adverb: 1.64; verb: 1.10). According to a 3×2 (Word class: mimetic words/verb/adverb × matched/mismatched) ANOVA, rating scores were significantly different between the matched and mismatched pairs ($F(1,10) = 166.06$, $p< 0.01$). The main effect of the word class was also significant ($F(1,10) = 5.53$, $p<0.05$). As indicated by the post-hoc test, the significant main effect for the word class was due to the difference between adverbs and verbs ($p = 0.05$, Bonferroni corrected), but the rating scores were similar between mimetic words and adverbs and between mimetic words and verbs ($p>0.05$). Reaction times (RTs) were not analyzed in Experiment 1 as participants were instructed to delay their response until each video clip was over.

Table 1. Activation for mimetic words (Experiment 1).

Region of activation	Lat.	Coordinates			T-score	*k*
		x	*y*	*z*		
(Matched motion-word pairs)						
superior temporal sulcus	R	52	−36	14	6.53	398
post-central gyrus	R	40	−20	44	9.63	145
parahippocampal gyrus	L	−30	−10	−14	7.19	170
Cerebellum	L	−28	−38	−36	7.18	151
(Mismatched motion-word pairs)						
parahippocampal gyrus	L	−26	−18	−18	5.98	251
inferior frontal gyrus	R	42	22	8	5.56	288

Note: coordinates (mm) are in MNI space. L = left hemisphere; R = right hemisphere. P<0.001 (uncorrected), *k*>140, P<0.05 (FWE corrected).
doi:10.1371/journal.pone.0097905.t001

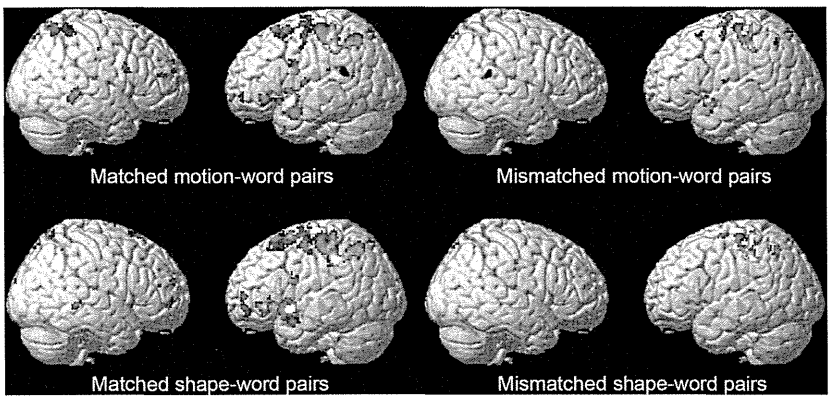

Figure 4. Results of Experiment 2. Several brain regions, including the right posterior STS, showed significant activation compared to baseline ($p<0.01$, FWE corrected). A strict threshold was used to ensure that the right pSTS was involved in the processing of both motion mimetic words and shape mimetic words. Brain activity observed during the motion and shape trials were not significantly different. However, the high-match trials elicited greater activation across the cortex than did low-match trials for either modality.
doi:10.1371/journal.pone.0097905.g004

Activation pattern for mimetic words. To identify the areas of activation for mimetic words, the images for verbs and adverbs were subtracted from the image of mimetic words (Figure 3; Table 1). As predicted, activation of the posterior part of the right STS was specific to mimetic words ($x = 52$, $y = -36$, $z = 14$; $T = 6.53$; $p<0.05$, FWE corrected). The post-central gyrus, parahippocampal gyrus, and cerebellum were also activated by mimetic words. In contrast, we confirmed the right posterior STS was not significantly activated when motion and mimetic words were mismatched ($p>0.05$, FWE corrected). Mismatched mimetic motion-word pairs showed increased activation in the parahippocampal gyrus and inferior temporal gyrus.

Experiment 2

Behavioral results. We examined whether Modality (motion or shape) and Degree of Match (high or low) affected RTs. Using a two-way repeated measures ANOVA, we found that the RTs were significantly longer for the motion-word pairs than for the shape-word pairs ($F(1,10) = 7.33$, $p = 0.02$). However, there was no effect of Degree of Match ($F(1,10) = 2.13$, $p = 0.18$), or interaction between Modality and Degree of Match ($F(1,10) = 0.88$, $p = 0.37$).

General neural activation. Several brain regions, including the right posterior STS, showed significant activation compared to baseline (Table S1; Figure 4). Importantly, brain activation observed in the motion and shape trials did not significantly differ. However, the high-match trials elicited greater activation across the cortex than did low-match trials for either modality.

Region of interest analysis of the right posterior STS. The right posterior STS was activated in both motion and shape trials. To investigate whether the right posterior STS was activated for all conditions (matched motion, mismatched motion, matched shape, and mismatched shape), we performed a region of interest (ROI) analysis with a 3-mm ROI located at $x = 62$, $y = -38$, $z = -2$ (Figure 5). This region was chosen based on the local maximum coordinates for the right posterior STS region. Although the right posterior STS was activated for all conditions compared to baseline, stronger activation was observed for matched motion/shape-word pairs. The main effect of Degree of Match was statistically significant (two-way ANOVA; $F(1,10) = 8.06$, $p = 0.02$); however, there was no significant effect of Modality ($F(1,10) = 1.61$, $p = 0.23$) or interaction between Modality and Degree of Match ($F(1,10) = 0.84$, $p = 0.38$). The RTs for the Degree of Match judgment were not different across matched and mismatched pairs. The behavioral results suggest that the task difficulty did not differ between the matched and mismatched pairs in which we found a difference in neural activation.

Discussion

In the present study, we investigated neural processing of sound symbolism using Japanese phenomimes. Despite accumulating evidence of universal sensitivity to sound symbolic word–meaning correspondences, the neural mechanism underlying this phenomenon has not been determined. Based on the idea of a functional dissociation of left and right STS [24], we hypothesized that the right posterior STS plays a critical role in sound symbolism processing. Experiment 1 tested this hypothesis by comparing Japanese mimetic words with verbs and adverbs for human motions. Supporting our hypothesis, mimetic words activated the right posterior STS even though all mimetic words were visually-presented phenomimes. This finding suggests that the function of

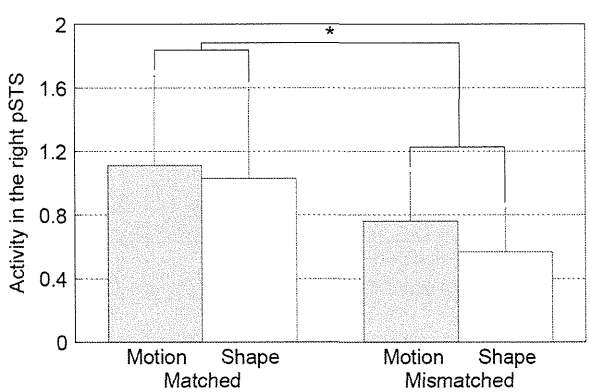

Figure 5. Mean beta values calculated in the ROI analysis for the right posterior STS (Experiment 2). Increased activation was observed for matched motion/shape-word pairs. A two-way ANOVA revealed the main effect of Degree of Match was statistically significant ($F(1,10) = 8.06$, $p = 0.02$). Error bars indicate ± 1 standard deviation.
doi:10.1371/journal.pone.0097905.g005

the right posterior STS is not limited to the processing of onomatopoeia; rather, this area is likely responsible for the processing of mimetic words.

The results of Experiment 1, however, might indicate an alternative interpretation. STS is known to have multiple functions, one of which is processing biological motion. Right posterior STS activation may reflect an enhanced response to human motion caused by the paired presentation of biological motion and a mimetic word. Experiment 2 ruled out this alternative as the right posterior STS was activated not only for motion mimetic words, but also for mimetic words representing static shapes. For both motion and shape trials, the level of right posterior STS activation was higher when the mimetic word and referent were highly matched, further confirming that this region is sensitive to sound symbolism. We thus conclude that the posterior STS serves as a critical hub for processing Japanese mimetic words, and possibly sound symbolism in general.

Previous neuroimaging research on the neural processing of Japanese mimetic words focused only on the acoustic similarity between onomatopoeia (phonomimes) and environmental sounds [25] or the "embodied" explanation of sound symbolism [35]. The embodied explanation of sound symbolism suggests that a mimetic word activates a perceptual or sensorimotor area relevant to the word meaning. Similar claims have also been made for non-mimetic conventional words [36–43]. Thus, the embodied explanation does not explain why people sense the meaning in the sound of the word. In contrast, we suggest that sound symbolism processing requires a unique neural basis involving the posterior STS.

Although we must be cautious about drawing a reverse inference, the unique involvement of the right posterior STS in mimetic processing supports the idea that sound symbolic words are processed as both linguistic symbols and non-linguistic iconic symbols. We speculate that the posterior STS works as a hub of multimodal integration. Our view, therefore, corroborates that of Ramachandran and Hubbard [4] that linked the neural mechanisms of sound symbolism to synesthesia. Event-related potential (ERP) studies also suggest that sensory integration at the parietal-occipital regions is related to sound symbolism processing [42].

Importantly, however, Ramachandran and Hubbard [4] identified the (left) angular gyrus as the key region for high-level synesthesia. The angular gyrus and the STS are closely located but are two distinct structures. The disparity between these two claims may suggest that the distinct neural substrates are involved in sound symbolism and synesthesia. Alternatively, the nature of stimuli types may have affected the results, as sound symbolic words differ greatly in terms of modalities and level of iconicity. Our stimuli consisted of sound symbolic words that are part of the Japanese lexicon rather than nonsense sound symbolic words (e.g., "baluma" and "takete"). Conventional words and nonsense sound symbolic words may in part recruit different processing mechanisms.

Future research is required to investigate other types of sound symbolic words including psychomimes, non-mimetic words that carry sound symbolism (e.g., English sound emission verbs), and non-lexicalized sound symbolic words (e.g., "baluma" and "takete"). Comparative imaging research that includes other populations, such as young children, non-Japanese speakers, or synesthetic individuals, would also improve understanding of the origin of sound symbolism. Although research on the neural mechanisms underlying sound symbolism is in its early stages, such research has the potential to advance our understanding of language.

Sound symbolism is not a marginal phenomenon in language. Developmental research has demonstrated that Japanese mothers often say mimetic words to their children [44], and sound symbolism of Japanese mimetic words promotes verb learning [18–20]. Sound symbolism may also play a key role in revealing the origin of language. Some researchers suggest that when human language started with our primitive ancestors, it began through their oral mimicking of the observed world [4,17]. Thus, mimetic words may be similar to the words our ancestors used as a form of protolanguage. Sound symbolism, as a bridge between non-speech sound and conventional words, can provide new insights into the ontogenesis and phylogenesis of language.

Supporting Information

Table S1 Cortical activation for figure and motion (Experiment 2).
(XLSX)

Acknowledgments

We thank Mamiko Arata for her contribution in stimuli development and data collection and Guillaume Thierry for his insightful comments on an earlier version of the manuscript.

Author Contributions

Conceived and designed the experiments: JK TM MI JO HO. Performed the experiments: JK TM. Analyzed the data: JK TM. Contributed reagents/materials/analysis tools: JK TM. Wrote the paper: JK TM MI.

References

1. de Saussure F (1983) Course in general linguistics. La Salle: Open Court. 236 p.
2. Jakobson R, Waugh L (1979) The sound shape of language. Bloomington: Indiana University Press. 335p.
3. Köhler W (1929) Gestalt psychology. New York: Liveright Publishing Corporation. 384 p.
4. Ramachandran VS, Hubbard EM (2001) Synesthesia– a window into perception, thought, and language. J Cons Stud 8: 3–34.
5. Hinton L, Nichols J, Ohala J, editors (1994) Sound Symbolism. Cambridge: Cambridge University Press.
6. Firth JR (1957) The use and distribution of certain English sounds. In: Firth JR, editor. Papers in linguistics 1934–1951. London: Oxford University Press. PP. 34–46.
7. Childs GT (1994) African ideophones. In: Hinton L, Nichols J, Ohala JJ, editors. Sound symbolism. Cambridge: Cambridge University Press. PP. 178–206.
8. Diffloth G (1972) Notes on expressive meaning. In: Peranteau PM, Levi JN, Phares GC, editors. Papers from the eighth regional meeting of Chicago linguistic society. Chicago: Chicago Linguistic Society. PP. 440–447.
9. Lee J-S (1992) Phonology and sound symbolism in Korean ideophones [dissertation], Bloomington: Indiana University.
10. Nuckolls JB (1999) The case for sound symbolism. Annu Rev Anthropol 28: 225–252.
11. Mok WE (2001) Chinese sound symbolism: A phonological perspective [dissertation], Honolulu: University of Hawaii.
12. Watson RL (2001) A comparison of some Southeast Asian ideophones with some African ideophones. In: Voeltz FKE, Kilian-Hatz C, editors. Ideophones. Amsterdam: John Benjamins. PP. 385–405.
13. Enfield NJ (2005) Areal linguistics and mainland Southeast Asia. Annu Rev Anthropol 34: 181–206.
14. Bodomo A (2006) The structure of ideophones in African and Asian languages: the case of Dagaare and Cantonese. In: Mugane J, editor. Selected proceedings of the 35th annual conference on African linguistics. Somerville: Cascadilla Proceedings Project. PP. 203–213.
15. Nygaard LC, Cook AE, Namy LL (2009) Sound to meaning correspondences facilitate word learning. Cognition 112: 181–186.
16. Kovic V, Plunkett K, Westermann G (2010) The shape of words in the brain. Cognition 114: 19–28.
17. Maurer D, Pathman T, Mondloch CJ (2006) The shape of boubas: Sound-shape correspondences in toddlers and adults. Dev Sci 9: 316–322.

18. Imai M, Kita S, Nagumo M, Okada H (2008) Sound symbolism facilitates early verb learning. Cognition 109: 54–65.
19. Kita S, Kantartzis K, Imai M (2010) Children learn sound symbolic words better: Evolutionary vestige of sound symbolic protolanguage. In: Smith ADM, Schouwstra M, de Boer B, Smith K, editors. Proceedings of the 8th conference of Evolution of Language. World Scientific. London. PP. 206–213.
20. Kantartzi K, Imai M, Kita S (2011) Japanese sound symbolism facilitates word learning in English speaking children. Cogn Sci 35: 575–586.
21. Pena M, Mehler J, Nespor M (2011) The role of audiovisual processing in early conceptual development. Psychol Sci 22: 1419–1421.
22. Ozturk O, Krehm M, Vouloumanos A (2013) Sound symbolism in infancy: Evidence for sound-shape cross-modal corespondences in 4-month-olds. J Exp Child Psychol 114: 173–186.
23. Arbib MA (2005) From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. Behav Brain Sci 28: 105–167.
24. Thierry G, Giraud AL, Price C (2003) Hemispheric dissociation in access to the human semantic system. Neuron 38: 499–506.
25. Hashimoto T, Usui N, Taira M, Nose I, Haji T, et al. (2006) The neural mechanism associated with the processing of onomatopoetic sounds. Neuroimage 31: 1762–1770.
26. Akita K (2009) A grammar of sound-symbolic words in Japanese: Theoretical approaches to iconic and lexical properties of mimetics [dissertation], Kobe: Kobe University.
27. Iwasaki N, Vinson DP, Vigliocco G (2007) What do English speakers know about gera-gera and yota-yota? A cross-linguistic investigation of mimetic words for laughing and walking. Japanese-Language Education around the Globe 17: 53–78.
28. Iwasaki N, Vinson DP, Vigliocco G (2007) How does it hurt, "kiri-kiri" or "siku-siku"? Japanese mimetic words of pain perceived by Japanese speakers and English speakers. In: Minami M, editor. Applying theory and research to learning Japanese as a foreign language. Newcastle upon Tyne: Cambridge Scholars Publishing. PP. 2–19.
29. Pelphrey KA, Singerman JD, Allison T, McCarthy G (2003) Brain activation evoked by perception of gaze shifts: the influence of context. Neuropsychologia 41: 156–170.
30. Thompson JC, Clarke M, Stewart T, Puce A (2005) Configural processing of biological motion in human superior temporal sulcus. J Neurosci 25: 9059–9066.
31. Grossman E, Blake R (2002) Brain Areas Active during Visual Perception of Biological Motion. Neuron 35: 1157–1165.
32. Holmes AP, Josephs O, Büchel C, Friston KJ (1997) Statistical modeling of low-frequency confounds in fMRI. Proceeding of the 3rd international conference of the functional mapping of the human brain. Neuroimage 5: S480.
33. Friston KJ, Josephs O, Zarahn E, Holmes AP, Rouquette S, et al. (2000) To smooth or not to smooth? Bias and efficiency in fMRI time-series analysis. Neuroimage 12: 196–208.
34. Ono M, editor (2007) Giongo/Gitaigo 4500 nihongo onomatope jiten [Japanese Onomatopoeia Dictionary 4500 mimetics]. Tokyo: Shogakukan. 770 p.
35. Osaka N, Osaka M, Morishita M, Kondo H, Fukuyama H (2004) A word expressing affective pain activates the anterior cingulate cortex in the human brain: an fMRI study. Behav Brain Res 153: 123–7.
36. Martin A, Ungerleider LG, Haxby JV (2000) Category-specificity and the brain: The sensory-motor model of semantic representations of objects. In Gazzaniga MS, editor. The new cognitive neurosciences. 2nd ed. Cambridge: MIT Press. PP. 1023–1036.
37. Pulvermüller F, Harle M, Hummel F (2001) Walking or talking? Behavioral and neurophysiological correlates of action verb processing. Brain Lang 78(2): 143–168.
38. Hauk O, Johnsrude I, Pulvermüller F (2004) Somatotopic representation of action words in human motor and premotor cortex. Neuron 41: 301–307.
39. Hauk O, Pulvermüller F (2004) Neurophysiological distinction of action words in the fronto-central cortex. Hum Brain Mapp 21: 191–201.
40. Kable JW, Kan I, Wilson A, Thompson-Schill S, Chatterjee A (2005) Conceptual representations of action in lateral temporal cortex. J Cogn Neurosci 17: 1855–1870.
41. Tettamanti M, Buccino G, Saccuman MC, Gallese V, Danna M, et al. (2005) Listening to action related sentences activates fronto-parietal motor circuits. J Cogn Neurosci 17(2): 273–281.
42. Aziz-Zadeh L, Wilson SM, Rizzolatti G, Iacoboni M (2006) Congruent embodied representations for visually presented actions and linguistic phrases describing actions. Curr Biol 16(18): 1818–1823.
43. Kemmerer D, Tranel D (2008) Searching for the elusive neural substrates of body part terms: a neuropsychological study. Cogn Neuropsychol 25: 601–29.
44. Fernald A, Morikawa H (1993) Common themes and cultural variations in Japanese and American mothers' speech to infants. Phonetica 57: 242–254.

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 24 May 2013
doi: 10.3389/fnhum.2013.00221

Cerebral responses to vocal attractiveness and auditory hallucinations in schizophrenia: a functional MRI study

Michihiko Koeda[1,2]*, Hidehiko Takahashi[3], Masato Matsuura[4], Kunihiko Asai[5] and Yoshiro Okubo[2]

[1] *Voice Neurocognition Laboratory, The Centre for Cognitive Neuroimaging, The Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, UK*
[2] *Department of Neuropsychiatry, Nippon Medical School, Tokyo, Japan*
[3] *Department of Psychiatry, Kyoto University, Kyoto, Japan*
[4] *Department of Biofunctional Informatics, Tokyo Medical and Dental University, Tokyo, Japan*
[5] *Asai Hospital, Chiba, Japan*

Edited by:
Paul Allen, King's College London, UK

Reviewed by:
Thomas P. White, King's College London, UK
Renaud Jardri, University Medical Centre of Lille, France

****Correspondence:***
Michihiko Koeda, Department of Neuropsychiatry, Nippon Medical School, 1-1-5 Sendagi, Bunkyo-ku, Tokyo 113-8603, Japan.
e-mail: mkoeda@nms.ac.jp

Impaired self-monitoring and abnormalities of cognitive bias have been implicated as cognitive mechanisms of hallucination; regions fundamental to these processes including inferior frontal gyrus (IFG) and superior temporal gyrus (STG) are abnormally activated in individuals that hallucinate. A recent study showed activation in IFG-STG to be modulated by auditory attractiveness, but no study has investigated whether these IFG-STG activations are impaired in schizophrenia. We aimed to clarify the cerebral function underlying the perception of auditory attractiveness in schizophrenia patients. Cerebral activation was examined in 18 schizophrenia patients and 18 controls when performing Favorability Judgment Task (FJT) and Gender Differentiation Task (GDT) for pairs of greetings using event-related functional MRI. A full-factorial analysis revealed that the main effect of task was associated with activation of left IFG and STG. The main effect of Group revealed less activation of left STG in schizophrenia compared with controls, whereas significantly greater activation in schizophrenia than in controls was revealed at the left middle frontal gyrus (MFG), right temporo-parietal junction (TPJ), right occipital lobe, and right amygdala ($p < 0.05$, FDR-corrected). A significant positive correlation was observed at the right TPJ and right MFG between cerebral activation under FJT minus GDT contrast and the score of hallucinatory behavior on the Positive and Negative Symptom Scale. Findings of hypo-activation in the left STG could designate brain dysfunction in accessing vocal attractiveness in schizophrenia, whereas hyper-activation in the right TPJ and MFG may reflect the process of mentalizing other person's behavior by auditory hallucination by abnormality of cognitive bias.

Keywords: attractiveness, auditory hallucinations, schizophrenia, greeting, cerebral laterality, social communications, functional MRI

INTRODUCTION

Auditory hallucinations and thought disorder are the main symptoms of schizophrenia, and these symptoms profoundly affect the neural basis of social communications as well as behavior (Brune et al., 2008; Bucci et al., 2008; Wible et al., 2009; Kumari et al., 2010; Granholm et al., 2012; Waters et al., 2012). In order to understand these psychiatric symptoms in schizophrenia, it is important to verify the pathophysiology of cerebral function in auditory communications.

For healthy people, greeting conversations are very essential tools for communicating socially with family, friends, and community. Since favorable greetings strengthen cordial relationships with colleagues, maintaining the skill of socializing with greeting conversations is especially significant (Gronna et al., 1999; Barry et al., 2003). One of the main cognitive models in schizophrenia proposes that hallucinations arise from impaired self-monitoring and abnormality of cognitive bias (Allen et al., 2004). Some studies indicate that schizophrenia patients tend to misapprehend inner speech as external speech by the disturbance of self-monitoring (Morrison and Haddock, 1997; Stein and Richardson, 1999; Ford et al., 2001; Allen et al., 2004). A recent study has suggested that auditory hallucination in schizophrenia may be caused by both impaired brain function in auditory processing and disturbance of attention bias toward internally generated information (Kompus et al., 2011). If patients with schizophrenia mistake unfavorable greetings through their distorted thinking while listening to favorable greetings, social isolation and emotional withdrawal could be produced. In addition, if schizophrenia patients have auditory hallucination, misjudgment of favorable/unfavorable greeting may be induced by abnormality of cognitive bias. However, it is unclear whether schizophrenia patients with auditory hallucinations have impaired abilities to differentiate between favorable and unfavorable greetings.

Functional magnetic resonance imaging (fMRI) studies in schizophrenia have investigated the neural basis of impairment of paralinguistic processing such as emotional prosody and affective vocalizations (Mitchell et al., 2004; Leitman et al., 2007, 2011; Bach et al., 2009; Dickey et al., 2010) as well as language

processing (Woodruff et al., 1997; Kircher et al., 2001; Mitchell et al., 2001; Sommer et al., 2001; Schettino et al., 2010). A previous fMRI study concerning the recognition of emotional speech prosody demonstrated that temporal activation in schizophrenia patients was predominant in the left hemisphere, whereas that in normal control subjects showed right hemispheric dominance (Mitchell et al., 2004). Another fMRI study also found right-lateralized activation in healthy controls in the temporal-parietal region while listening to emotional prosody including meaningless syllables (Bach et al., 2009). In schizophrenia patients, however, this right-lateralized pattern was even more pronounced. These findings indicate that cerebral laterality for emotional prosody in schizophrenia patients could be shifted in comparison to the typical right-lateralized activation in normal control subjects.

We consider that it is important to investigate the relationship between psychiatric symptom and cerebral function in behavior social as well as emotional prosody. Especially, evaluating facial attractiveness is a favorable behavior associated with social communication (Kampe et al., 2001; Winston et al., 2007). Recent studies have demonstrated that facial attractiveness can activate doperminergic regions including amygdala and orbitofrontal cortex that are strongly related to reward prediction (Winston et al., 2007; Cloutier et al., 2008; Chatterjee et al., 2009; Tsukiura and Cabeza, 2011). Clarifying brain mechanisms in these reward systems is very important for understanding the pathophysiology of schizophrenia. A recent study has shown that in schizophrenia, the ratings of attractiveness of unfamiliar faces were significantly reduced compared to healthy subjects (Haut and MacDonald, 2010). Further, this study has demonstrated that when the patient had severe persecutory delusions, attractiveness ratings decreased (Haut and MacDonald, 2010). As well as facial perception, auditory attractiveness in schizophrenia will be a challenging research topic. A recent fMRI study in healthy subjects on auditory attractiveness has demonstrated bilateral superior temporal gyrus (STG) and inferior frontal gyrus (IFG) activates when participants judged whether voices sounded attractive or not. This study suggests that the roles of STG and IFG are essential for perceiving auditory attractiveness (Bestelmeyer et al., 2012). The regions of STG and IFG are heavily implicated in the functional anatomy of auditory hallucination. A recent meta-analysis demonstrated that schizophrenia patients with auditory hallucination had significantly increased activity in fronto-temporal areas involved in speech generation and speech perception (Jardri et al., 2011). A recent fMRI study demonstrated that cerebral activation in fronto-temporal regions is greater than in healthy individuals during AVH but lower during environmental-stimulus processing (Kompus et al., 2011). However, to our knowledge, no study has ever investigated the cerebral response to auditory attractiveness in schizophrenia.

The aim of our research is to clarify cerebral response to auditory attractiveness when patients with schizophrenia are listening to greetings. Greeting conversations are crucial to maintaining social interactions. An fMRI study of social perception indicated that the left prefrontal and left IFG were activated when the subjects judged whether two people were friends or enemies (Farrow et al., 2011). Since the recognition of friendliness and favorability is essential for greeting conversations, the patients with schizophrenia could change cerebral function due to psychiatric symptoms such as auditory hallucinations. To investigate this pathophysiology, using completely the same greetings, we compared cerebral activation when the subjects judged favorability (recognition of auditory attractiveness) and cerebral activation when the subjects judged gender (recognition of non-auditory attractiveness). Prior to the current experiment, we hypothesized that cerebral functions underlying the perception of auditory attractiveness could be impaired in STG and IFG by occurring auditory hallucination.

MATERIALS AND METHODS

SUBJECTS OF fMRI STUDY

Eighteen right-handed controls (9 males and 9 females, mean age 35.5 years, $SD = 8.6$) and 18 schizophrenia patients (10 males and 8 females, mean age 35.7 years, $SD = 8.4$) participated in the present study. As for the subtypes of 18 schizophrenia patients, all patients were diagnosed with paranoid schizophrenia. All 18 patients were receiving neuroleptics (mean risperidone equivalent daily dosage = 4.7 mg, $SD = 2.2$; 9 patients, risperidone; 4 patients, olanzapine; 2 patients, haloperidol; 1 patient, quetiapine; 1 patient, sulpiride; 1 patient, perphenazine). Risperidone equivalents were calculated based on published equivalencies for atypical antipsychotics by Inagaki and Inada (2006). All 36 volunteers were native speakers of Japanese. None of the control subjects was taking alcohol or medication at the time, nor did they have a history of psychiatric disorder, significant physical illness, head injury, neurological disorder, or alcohol or drug dependence. After complete explanation of the study, written informed consent was obtained from all subjects, and the study was approved by the relevant ethics committee. Schizophrenia patients were diagnosed by MK and the attending psychiatrists on the basis of a review of their charts and a conventionally semi-structured interview (First et al., 1995). After the structural interview was performed using PANSS, the patient was synthetically diagnosed according to the diagnostic guidelines of the ICD-10: Classification of Mental and Behavioral Disorders. Exclusion criteria were current or past substance abuse and a history of alcohol-related problems, mood disorder, or organic brain disease. All patients were recruited from the outpatient unit of Asai Hospital. Mean illness duration was 12.3 ($SD = 8.0$) years. Clinical symptoms were assessed by Positive and Negative Syndrome Scale (PANSS) (Kay et al., 1987). Sum scores for positive and negative symptoms were calculated, with the positive symptom subscale including the following seven items: Delusion, Conceptual disorientation, Hallucinatory behavior, Excitement, Grandiosity, Suspiciousness, and Hostility. The negative symptom subscale also included seven items: Blunted affect, Emotional withdrawal, Poor rapport, Passive/apathetic social withdrawal, Difficulty in abstract thinking, Lack of spontaneity and flow of conversation, and Stereotyped thinking. The mean score of PANSS was 32.4 ($SD = 10.4$). The mean positive symptom score was 15.1 ($SD = 6.4$), mean negative symptom score was 20.7 ($SD = 6.2$), and mean score of general psychopathology was 32.3 ($SD = 7.9$). The candidates were carefully screened and standardized interviews were conducted by a research psychiatrist (MK)

and the attending psychiatrists. They did not meet the criteria for any psychiatric disorders. There was no significant difference in the mean period of education between the controls and patients (mean ± *SD*; patients 13.3 ± 1.3 years, control subjects 13.0 ± 1.0 years; $p > 0.05$, *t*-test). Schizophrenia patients were 14 right-handed and 4 left-handed participants according to the Edinburgh Handedness Inventory (EHI) (Oldfield, 1971). Mean (± *SD*) EHI in right-handed 14 patients was 90.4 ± 13.0. The EHI score of the 4 left-handed patients was −85, −73, −46, −46, respectively. All control subjects were right-handed, and mean (± *SD*) EHI was 96.1 ± 4.7.

RECORDED VOICE

As a sample for clarifying emotional response in voice recognition, Japanese greetings were recorded from 6 native speakers (3 males, 3 females). Ten greetings were recorded: Ohayo (Good Morning), Yah (Hi), Konnichiwa (Good Afternoon), Konbanwa (Good evening), Arigato (Thank you), Domo (Thank you), Irasshai (Welcome), Genki (How are you?), Dozo (Please), and Hisashiburi (Long time no see). These 10 greetings were recorded expressing favorable emotion (positive greeting), unfavorable emotion (negative greeting), or without emotion (neutral greeting), resulting in 180 stimuli in total. The voice was recorded using an IC recorder (Voice-Trek DS-71, Olympus) in a perfectly quiet room. In both the preliminary experiment and the fMRI experiment, all speakers were unknown to all participants.

PRELIMINARY EXAMINATION

Prior to the fMRI study, we asked 32 different control volunteers (16 males and 16 females) to judge the favorability of all 180 greetings (60 favorable, 60 non-favorable, and 60 neutral greetings) using a questionnaire with a 10-point scale. We defined "favorable" if the scale approached 10, whereas "unfavorable" if the scale approached 0. Based on the responses of the 32 subjects, greetings were considered positive if their average score was higher than 6.5. If the average score was less than 3.5, the greetings were considered negative. Neutral greetings were defined as being located within the average score range of 4.5–5.5. Based on these results, each speaker's greeting was evenly selected for the favorable, neutral, and unfavorable greetings.

INSTRUMENTS USED FOR PRESENTATION OF STIMULI

Stimuli were presented by the use of Media Studio Pro (version 6.0 Ulead Systems, Inc., Ulead Systems, Taiwan) running under Windows XP. Subjects listened to the sound stimuli through headphones attached to an air conductance sound delivery system (Commancer X6, MRI Audio System, Resonance Technology Inc., Los Angeles, CA). The average sound pressure of stimulus amplitude was kept at 80 dB.

EXPERIMENT DESIGN

The subjects listened for a total of 10 min and 40 s: 20 s of silence, 5 min of attentive listening (Part A), 20 s of silence, and 5 min of attentive listening (Part B). Part A and Part B each consisted of 60 paired greetings (30: neutral-positive, 30 neutral-negative), with each greeting taking 0.5 s, and pause of 1 s; all together, each of the 10 greetings was spoken 12 times (6 times as a neutral greeting, 3 times as a positive greeting, and 3 times as a negative greeting). In Part A, there were equal numbers of greetings by male pairs and female pairs. The subjects judged which greeting of a pair was more favorable. Using 30 neutral-positive and 30 neutral-negative pairs, we examined the degree of difference in favorability. We named Part A: Favorability Judgment Task (FJT). In Part B, 30 pairs were the same gender and 30 pairs were different gender. The subjects judged whether the speakers in each pair were the same gender or not. We named Part B: Gender Discrimination Task (GDT). The pairings of neutral-positive and neutral-negative appeared in random order (**Figure 1**).

FUNCTIONAL MRI ACQUISITION

The images were acquired with a 1.5 Tesla Signa system (General Electric, Milwaukee, Wisconsin). Functional images of 264 volumes were acquired with T2*-weighted gradient echo planar imaging sequences sensitive to blood oxygenation level dependent (BOLD) contrast. Each volume consisted of 20 transaxial contiguous slices with a slice thickness of 6 mm to cover almost the whole brain (flip angle, 90°; time to echo [TE], 50 ms; repetition time [TR], 2.5 s; matrix, 64 × 64; field of view, 24 × 24).

IMAGE PROCESSING

Data analysis was performed with statistical parametric mapping software SPM8 (Wellcome Department of Cognitive Neurology, London, United Kingdom) running with MATLAB (Mathworks, Natick, Massachusetts). All volumes of functional EPI images were realigned to the first volume of each session to correct for subject motion, and the mean functional EPI image was spatially coregistered with the anatomical T1 images. The anatomical T1 image was segmented into the image of gray matter and white matter. Based on the segmented T1 image of each subject, the anatomical template of diffeomorphic anatomical registration through an exponentiated Lie algebra (DARTEL) was created (Ashburner, 2007). All realigned EPI images were spatially normalized to the standard space defined by the Montreal Neurological Institute (MNI) template with DARTEL template and flow field of each subject. Functional images were spatially smoothed with a 3-D isotropic Gaussian kernel (full width at half maximum of 8 mm). A temporal smoothing function was applied to the fMRI time series to enhance the temporal signal-to-noise ratio. The significance of hemodynamic changes in each condition was examined using the general linear model with boxcar functions convoluted with a hemodynamic response function. The *t*-values were then transformed to unit normal distribution, resulting in z-scores. The models of 4 contrasts were created by event-related design during the fMRI experiments. In FJT task, 2 contrasts [30 pairs of neutral-favorable greetings (FAV) and 30 pairs of neutral-unfavorable greetings (NFV)] were made. In GDT task, 2 contrasts [30 pairs of same gender greetings (SAM) and 30 pairs of different gender greetings (DIF)] were made (**Figure 1**).

STATISTICAL ANALYSIS

Group analysis (2nd-level analysis in spm8) was performed on the data for 18 control subjects and 18 schizophrenia patients using a random effect model on a voxel-by-voxel basis. FMRI data

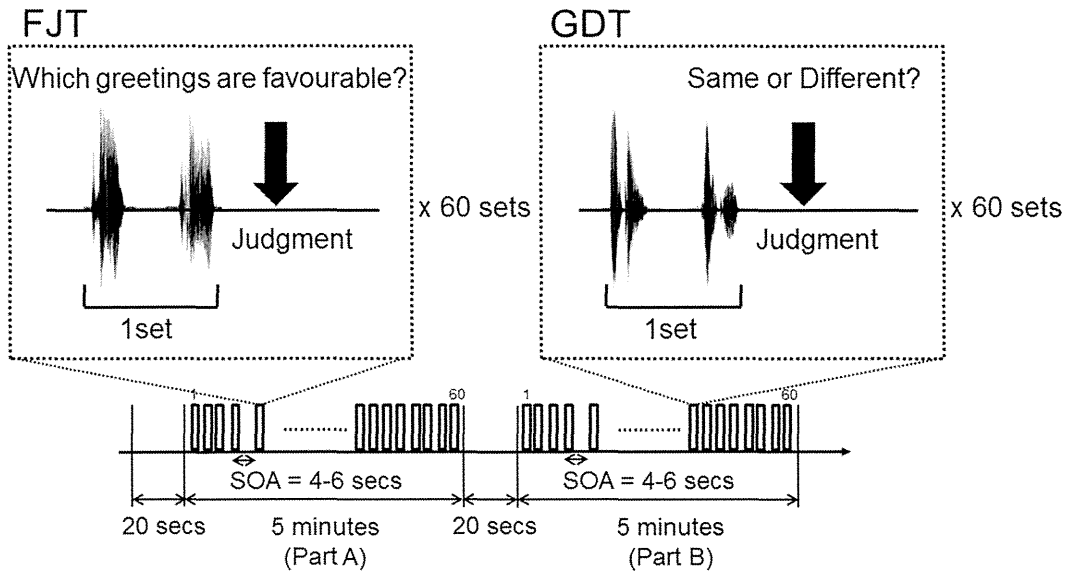

FIGURE 1 | Experimental design of fMRI Part A is Favorability Judgment Task (FJT). Part B is Gender Discrimination Task (GDT).

was analysed based on the 2 × 2 × 2 full factorial model with the factors of Group (control subjects/schizophrenia patients), Task (FJT/GDT) and Within-task (FJT: FAV/NFV, GDT: SAM/DIF) (FDR-corrected voxel-level threshold of $P < 0.05$). By using rfx-plot (Glascher, 2009), cerebral activation at the regions of interests (ROIs) was investigated. In main effect of Group and main effect of Task, ROIs were focused on the coordinates of the peak voxel of activation under FDR-corrected voxel-level threshold of $P < 0.05$. For main effect of Group, cerebral laterality of ROIs was evaluated. In order to investigate cerebral laterality, ROIs were also set on the hemispheric symmetrical region of the MNI coordinates. By using these symmetrical ROIs, the laterality index (LI) was calculated [LI = (L − R)/(L + R) × 100; L = beta estimates of left hemispheric activation, R = beta estimates of right hemispheric activation]. The formula of the LI was calculated based on previous studies (Koeda et al., 2006, 2007; White et al., 2009). In calculation of LI, the beta value of each subject used was either plus or zero, and minus beta values were excluded. In this ROI analysis, correlation between EHI score and beta value was evaluated to investigate the influence of handedness. Correlations between the subscores of PANSS (total scores of positive symptoms, negative symptoms, and general psychopathology) and cerebral activation under FJT vs. GDT contrast were calculated based on simple regression in schizophrenia patients. In linear regression analyses, the three subscores of PANSS were used, each with one predictor, respectively. In the analysis of full factorial design, the statistical threshold used was $p < 0.05$, voxel level, FDR-corrected. In the linear regression analysis, the statistical threshold used was $p < 0.0001$, voxel level, uncorrected (FDR < 0.25, voxel level corrected). Further, the correlation was analysed between the beta value of FJT at the specific ROIs of main effect of Group (**Figure 10**).

RESULTS

PRELIMINARY EXPERIMENTS

Favorability was rated by 32 different control volunteers using a scale of 1–10. **Figure 2** shows the distribution of the rating of favorability. Based on the definition of favorability (Materials and Methods: Preliminary Examination), 30 favorable vocalizations (rating average more than 6.5; 12 males and 9 females), 60 neutral vocalizations (rating average between 4.5 and 5.5; 17 males and 18 females), 30 unfavorable vocalizations (rating average less than 3.5; 7 males and 13 females) were selected. The mean ratings (± *SD*) of favorability were 2.3 ± 0.6 (unfavorable), 4.9 ± 0.3 (neutral), and 7.5 ± 0.6, respectively. Analysis of variance (One-Way ANOVA) was significantly different [$F_{(2,\ 117)} = 1006.9$, $p < 0.001$]. Multiple comparisons were also significant (unfavorable vs. neutral: 2.6 ± 0.1, $p < 0.001$; neutral vs. favorable: 2.6 ± 0.1, $p < 0.001$; unfavorable vs. favorable: 5.2 ± 0.1, $p < 0.001$).

Behavioral data (accuracy)

In the fMRI experiment, the mean percentages (±*SD*) of the accuracy of the control subjects for FJT and GDT were 94.7 ± 6.1% and 97.0 ± 3.1%, and those of schizophrenia patients were 90.9 ± 6.0% and 95.1 ± 4.8%, respectively (**Figure 3**). There was no significant difference between the two groups [FJT: $t_{(34)} = 1.89$, $p > 0.05$; GDT: $t_{(34)} = 1.45$, $p > 0.05$]. Mixed analysis of variance (mixed ANOVA) in the performance did not show a significant main effect of Group (control subjects/schizophrenia patients): $F_{(1,\ 34)} = 2.33$, $p > 0.05$, whereas a significant Task effect (FJT vs. GDT) was observed: $F_{(1,\ 34)} = 11.5$, $p < 0.001$. No interaction effect between Group and Task was observed: $F_{(1,\ 34)} = 0.19$, $p > 0.05$. **Table 1** shows the mean accuracy for judgment of favorable/non-favorable, and judgment of same

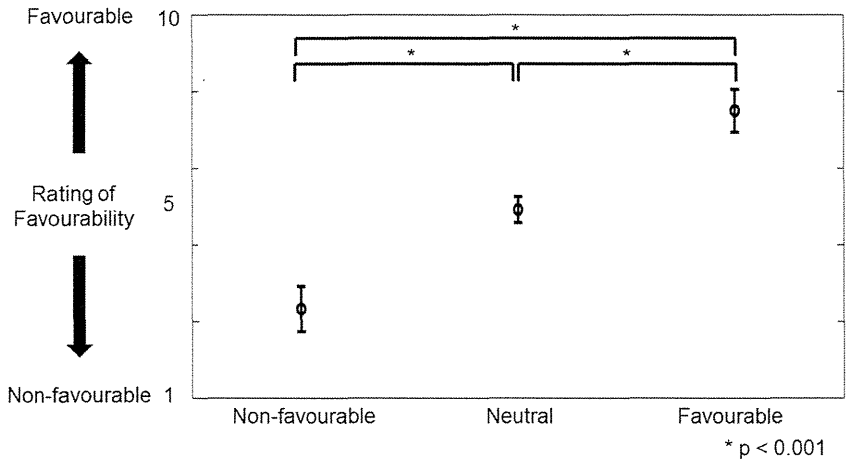

FIGURE 2 | The figure shows the distribution (mean ± *SD*) of rating favorability for portraying favorable, neutral, and unfavorable greetings in preliminary experiment. The vertical axis represents the rating score of favorability (1–10). *Indicates that multiple comparison was significant.

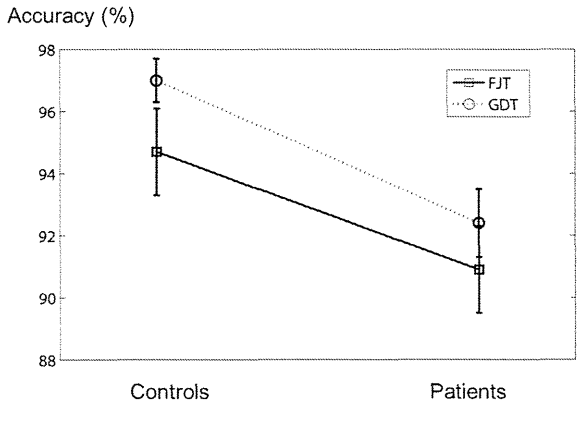

FIGURE 3 | The figure shows the error bar (mean ± *SD*) of the accuracy in the fMRI experiment. The vertical axis represents the accuracy of the experiment.

gender and different gender (FAV controls: 93.5 ± 7.6%; FAV patients: 94.1 ± 6.2%; NFV controls: 94.7 ± 6.1%; NFV patients: 90.0 ± 6.0%; SAM controls: 98.3 ± 2.4%; SAM patients: 95.7 ± 5.6%; DIF controls: 97.2 ± 3.8%; DIF patients: 95.0 ± 4.2%). No significant difference was observed between controls and patients [FAV: $t_{(34)} = -0.24$, $p > 0.05$; NFV: $t_{(34)} = 1.89$, $p > 0.05$; SAM: $t_{(34)} = 1.82$, $p > 0.05$; DIF: $t_{(34)} = 1.67$, $p > 0.05$]. Three-Way ANOVA was calculated for the effect of Group, Task, and Within-task. Task effect was significantly observed [$F_{(1, 34)} = 11.9$, $p = 0.002$], whereas Group effect and Within-task effect were not observed [Group: $F_{(1, 34)} = 2.33$, $p > 0.05$; Within-task: $F_{(1, 34)} = 2.55$, $p > 0.05$]. Interaction effect was not observed in the effect of Group × Task [$F_{(1, 34)} = 0.17$, $p > 0.05$] and the effect of Group × Within-task [$F_{(1, 34)} = 2.80$, $p > 0.05$].

Response time

The mean (± *SD*) response times relative to offset of stimulus (seconds) of control subjects and schizophrenia patients for FJT and GDT were the following: FJT-control: 2.41 ± 0.78 s, FJT-patients: 2.01 ± 0.33 s, GDT-control: 2.15 ± 0.78 s, and GDT-patients 1.99 ± 0.05 s. The response times in FJT were significantly different between control subjects and schizophrenia patients: $t_{(34)} = 2.16$, $p = 0.04 < 0.05$, whereas there was no significant difference in the response times to GDT: $t_{(34)} = 0.84$ $p > 0.05$. Analysis of variance on the response time was performed with the factors of Group (control subjects/schizophrenia patients) and Task (FJT/GDT). There was no significant difference between Groups: $F_{(1, 34)} = 2.29$, $p > 0.05$, whereas there was a significant Task effect: $F_{(1, 34)} = 33.7$, $p < 0.001$. An interaction effect between Group and Task was also observed: $F_{(1, 34)} = 24.4$, $p < 0.001$. **Table 1** (lower part) shows the mean response time for judgment of favorable/non-favorable, and judgment of same gender and different gender [FAV controls: 2.41 ± 0.78 s; FAV patients: 2.17 ± 0.17 s; NFV controls: 2.01 ± 0.13 s; NFV patients: 1.89 ± 0.15 s; SAM controls: 2.15 ± 0.78 s; SAM patients: 1.93 ± 0.14 s; DIF controls: 1.99 ± 0.53 s; DIF patients: 2.07 ± 0.13 s]. Significant difference between controls and patients was observed in NFV and DIF [NFV: $t_{(18.4)} = 3.49$, $p > 0.05$; DIF: $t_{(22.8)} = -2.31$, $p < 0.05$: Welch's t-test], whereas no significant difference was observed in FAV and SAM [FAV: $t_{(18.6)} = 1.29$, $p > 0.05$; SAM: $t_{(18.1)} = 1.15$, $p > 0.05$: Welch's t-test]. Three-Way ANOVA was calculated for the effect of Group, Task, and Within-task. Task effect was significantly observed [$F_{(1, 34)} = 36.7$, $p < 0.001$], whereas Group effect and Within-task effect were not observed [Group: $F_{(1, 34)} = 1.93$, $p > 0.05$; Within-task: $F_{(1, 34)} = 0.13$, $p > 0.05$]. Interaction effect was significantly observed in the effect of Group × Task [$F_{(1, 34)} = 16.0$, $p < 0.001$], whereas interaction effect was not significantly observed in the effect of Group × Within-task [$F_{(1, 34)} = 1.04$, $p > 0.05$].

Table 1 | Shows the mean ± *SD* of accuracy and response time in fMRI experiments.

		FAV	NFV	SAM	DIF
Accuracy (%)	Controls	93.5 ± 7.6	94.7 ± 6.1	98.3 ± 2.4	97.2 ± 3.8
	Patients	94.1 ± 6.2	90.9 ± 6.0	95.7 ± 5.6	95.0 ± 4.2
Response time (sec)	Controls	2.41 ± 0.78	2.01 ± 0.13	2.15 ± 0.78	1.99 ± 0.53
	Patients	2.17 ± 0.17	1.89 ± 0.15	1.93 ± 0.14	2.07 ± 0.13

FAV, pairs of neutral-favorable greetings; NFV, pairs of neutral-unfavorable greetings; SAM, pairs of same gender greetings; DIF, pairs of different gender greetings.

FUNCTIONAL MRI DATA

Full factorial design analysis

FMRI data was analysed based on the $2 \times 2 \times 2$ full factorial model with the three factors: Group (control subjects/schizophrenia patients), Task (FJT/GDT), and Within-task (FJT: FAV/NFV, GDT: SAM/DIF) (FDR-corrected voxel-level threshold of $P < 0.05$).

Main effect of Group was significantly observed in the bilateral middle frontal gyrus (MFG), left STG, right superior parietal lobe (SPL) temporo-parietal junction (TPJ), right occipital lobe, and right amygdala ($p < 0.05$, FDR-corrected, **Figure 4** and **Table 2**). The upper part (gray bar) of **Figure 4** shows the bar graph for contrast estimates and 90% confidence interval in each activated region (gray bar: controls, the light gray bar: patients). From the results of main effect of Group, ROIs were set on the 5 regions: left MFG [−26, −3, 62], right amygdala [20, −3, 21], left STG [−54, −21, 3], right TPJ [26, −65, 53], and right occipital lobe [8, −77, 2]. In these ROIs, Mann–Whitney test was calculated for beta values between controls and patients. The P-threshold was Bonferroni-corrected based on 5 tests being conducted. Cerebral activation in left STG was significantly greater in control subjects than in schizophrenia patients (L STG: $z = 3.10$, $p = 0.001 < 0.05/5$), whereas cerebral activations in the other regions were significantly greater in schizophrenia patients than in control subjects (L MFG: $z = -3.61$, $p < 0.05/5$; R amygdala: $z = -3.54$, $p < 0.05/5$; R TPJ: $z = -3.61$, $p < 0.05/5$; R occipital: $z = -3.54$, $p < 0.05/5$). In these 5 ROIs and contralateral symmetrical 5 ROIs (right MFG [26, −3, 62], left amygdala [−20, −3, 21], right STG [54, −21, 3], left TPJ [−26, −65, 53], left occipital lobe [−8, −77, 2]), cerebral activation under FAV, NFV, SAM, and DIF conditions was evaluated (middle part of **Figure 4**). Further, the LI was calculated (lower part of **Figure 4**). For each ROI, Two-Way ANOVA was calculated by main effect of Group and Within-task. Regarding Group effect, in the ROIs at the bilateral MFG, bilateral amygdala, bilateral TPJ, and right occipital lobe, the strength of BOLD signal (beta estimates) in patients under the FAV and NFV conditions was significantly greater than that in controls [L MFG: $F_{(1, 34)} = 14.1, p < 0.001$; R MFG: $F_{(1, 34)} = 21.5, p < 0.001$; L amygdala: $F_{(1, 34)} = 14.9, p < 0.001$; R amygdala: $F_{(1, 34)} = 18.0$, $p < 0.001$; L TPJ: $F_{(1, 34)} = 12.7$, $p < 0.001$; R TPJ: $F_{(1, 34)} = 4.67$, $p < 0.05$; R occipital: $F_{(1, 34)} = 14.4, p < 0.001$], whereas that in bilateral STG was significantly greater in controls than in patients [L STG: $F_{(1, 34)} = 12.7$, $p < 0.001$; R STG: $F_{(1, 34)} = 4.7$, $p < 0.05$]. In bilateral MFG, right amygdala, and right occipital, BOLD signals of patients under SAM and DIF conditions were significantly greater than in controls [L MFG: $F_{(1, 34)} = 14.1$, $p < 0.001$; R MFG: $F_{(1, 34)} = 21.5$, $p < 0.001$; R amygdala: $F_{(1, 34)} = 7.8$, $p < 0.01$; R occipital: $F_{(1, 34)} = 5.9$, $p < 0.05$]. Significant difference of LI was observed in the amygdala and occipital lobe under SAM and DIF conditions [amygdala LI: $F_{(1, 34)} = 7.8$, $p < 0.001$; occipital lobe: $F_{(1, 34)} = 6.5$, $p < 0.05$], whereas significant difference in the other regions was not observed ($p > 0.05$).

Main effect of Task (FJT/GDT) was significantly observed in the left precentral gyrus (PrCG), left MFG, left IFG, right insula, bilateral STG, left claustrum, and left cerebellum ($p < 0.05$, FDR-corrected, **Figure 5** and **Table 3**). Cerebral activation in the left IFG and bilateral STG was significantly greater in FJT than in GDT [**Figure 5**; L IFG: $t_{(70)} = 3.92$, $p < 0.05/6$; L STG: $t_{(70)} = 4.64$, $p < 0.05/6$; R STG: $t_{(70)} = 2.92$, $p = 0.005 < 0.05/6$]. Interaction effect between Group and Task was not significantly observed at a threshold of $p < 0.05$, FDR-corrected.

CORRELATION BETWEEN PSYCHIATRIC SYMPTOM AND CEREBRAL ACTIVATION

We examined correlations between PANSS and cerebral activation under FJT minus GDT contrast. Significant positive correlations were observed in the right superior frontal gyrus (SFG), right MFG, left IFG, left STG, and right IPL in schizophrenia ($p < 0.25$, FDR-corrected, **Figure 6** and **Table 4**). **Figure 7** demonstrated correlations between the severity of auditory hallucinations and cerebral activation under FJT minus GDT contrast. Significant positive correlations were observed in the right post central gyrus (PsCG), right PrCG, right MFG and right IPL in schizophrenia ($p < 0.25$, FDR-corrected, **Figure 7** and **Table 5**).

CORRELATION BETWEEN HANDEDNESS AND CEREBRAL ACTIVATION

We examined the correlation between handedness and cerebral activation. The beta value of ROI analysis in main effect of Group was used in this analysis. Cerebral activations in most ROIs were not correlated with the handedness score, but activation in STG was significantly negatively correlated with the handedness score (**Figure 8**). In the ROI of STG, differences in LI between 18 controls and 14 patients were analysed after removing 4 left-handed patients. However, significant difference in the LI was not observed.

CORRELATION BETWEEN ACCURACY OF TASK AND CEREBRAL ACTIVATION

In the ROIs of main effect of Group, correlation was analysed between the beta value of FJT at left STG and accuracy.