bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Title: Heartfelt Face Perception via the Interoceptive Pathway – a MEG study

Author Affiliation: Jaejoong Kim[1], Hyeong-Dong Park[2], Ko Woon Kim[1,3], Dong Woo Shin[1], Sanghyun Lim[4], Hyukchan Kwon[4], Min-Young Kim[4], Kiwoong Kim[4], Bumseok Jeong[1*]

[1]Computational Affective Neuroscience and Development Laboratory, KAIST, Graduate School of Medical Science and Engineering, Daejeon, Korea; KAIST Institute for Health Science and Technology, KAIST, Daejeon, Korea; [2]Laboratory of Cognitive Neuroscience, Center for Neuroprosthetics and Brain Mind Institute, Ecole Polytechnique Fédérale de Lausanne (EPFL), 9 Chemin des Mines, 1202 Geneva, Switzerland; [3]Department of Neurology, Chonbuk National University Hospital, Chonbuk National University Medical school, JeonJu, Korea; [4]Center for Biosignals, Korea Research Institute of Standards and Science, Daejeon, South Korea

Correspondence: Bumseok Jeong, M.D., Ph.D. Associate Professor, Director, Laboratory of Computational Affective Neuroscience and Development, Graduate School of Medical Science and Engineering, Korea Advanced Institute of Science and Technology, Daehak-ro 291, Daejeon, Korea

Tel. +82-42-350-4245 (Office), **E-mail:** bs.jeong@kaist.ac.kr

Keywords: Heartbeat-evoked potential, Interoception, Emotion, Magnetoencephalography, Granger causality

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

The somatic marker hypothesis proposes that cortical representation of visceral signals is a crucial component of emotion processing. No previous study has investigated the causal relationship among brain regions of visceral information processing during emotional perception. In this magnetoencephalography (MEG) study of 32 healthy subjects, heartbeat evoked potential (HEP), which reflects the cortical processing of heartbeats, was modulated by the perception of emotional faces. The modulation effect was localized in the prefrontal cortices, globus pallidus, and interoceptive network including right anterior insula (RAI) and anterior cingulate cortex (RACC). Importantly, our Granger Causality analysis provide the first evidence for an increased causal flow of heartbeat information from the RAI to the RACC by sad faces. Moreover, this HEP modulation effect was neither an artifact nor an effect of visual evoked potential. These findings provide an important progress in the understanding of brain-body interaction during emotion processing.

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

According to the James-Lange theory and the somatic marker hypothesis, emotional feelings are the mental experience of bodily states (A. Damasio & Carvalho, 2013; James, 1884). More specifically, emotional stimuli usually induce a change of bodily status (Williams et al., 2005). Then various feelings subsequently emerge from the perception of bodily status including a sense of viscera (A. Damasio & Carvalho, 2013). Many previous studies showed evidence of emotional stimulus-evoked somatic response. For example, a fearful stimulus enhances sympathetic responses like heart rate elevation and skin conductance modulation (Critchley et al., 2005; Williams et al., 2005), and a disgusting stimulus induces tachygastria (N. A. Harrison, Gray, Gianaros, & Critchley, 2010). Moreover, neuroimaging studies using fMRI or EEG have shown that generation of bodily responses by an emotional stimulus is related to activation of subcortical regions like the amygdala and hypothalamus (A. R. Damasio et al., 2000). On the other hand, there is rare evidence that supports a change in cortical interoceptive processing in the brain during the experience of emotional feeling. Given that signals of internal organs cannot be identified without explicit measuring devices like an electrocardiogram (ECG), it is difficult to investigate which brain activity was directly evoked by interoceptive signals. Thus, previous studies explored changes in the interoceptive processing during just the experience of emotional states and reported their relation to the anterior insula and anterior cingulate cortex (Adolfi et al., 2016; Critchley et al., 2005). Heartbeat evoked potential (HEP), which is obtained by averaging electrophysiological signals time-locked to heartbeats, has been reported to be associated with the perception of heartbeat (Pollatos & Schandry, 2004), pain (Shao, Shen, Wilder-Smith, & Li, 2011), and feeling empathy (Fukushima, Terasawa, & Umeda, 2011). Moreover, HEP amplitude is attenuated in mood-related psychiatric disorders, including depression (Terhaar, Viola, Bär, & Debener, 2012) and borderline personality disorder (Müller et al., 2015), suggesting a potential link between HEP and aberrant emotional processing.

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Based on these theories and indirect evidence, we hypothesized that HEP would be associated with the feeling of emotional stimuli and would be modulated during perception of emotional expression, such as a face or a text-based emoticon. And the modulation effect would be different between positive and negative valences. To test this idea, we used emotional faces and emotional emoticons that convey text-based emotion to evoke emotional feeling while measuring the HEP with magnetoencephalography (MEG). To verify the precise source of HEP modulation, T1-weighted structural magnetic resonance imaging (MRI) was acquired from all subjects. Importantly, we applied Granger causality analysis (Barnett & Seth, 2014) on sources to identify information flow between the sources of HEP modulation.

We formulated the following specific hypotheses. First, we expected that HEP would be modulated by the emotional expression and this effect would appear in different spatiotemporal dynamics between emotional and neutral stimulus presentation. Second, the modulation of HEP by emotional expression would be localized in the previously known interoceptive region, like the anterior insula and the anterior cingulate cortex, in source level analysis. Third, we expected that information flow between these interoceptive regions would be modulated by emotional expression. To be more specific, we expected that bottom up heartbeat information processing starting from the anterior insula, which re-represent the viscerosensory information from the posterior insula, to the anterior cingulate cortex would be enhanced by emotional expression (Medford & Critchley, 2010). This pathway is hypothesized to be involved in generating emotional feeling (Medford & Critchley, 2010; Smith & Lane, 2015). Finally, we hypothesized that neural activity that are evoked by heartbeat would have spatiotemporal patterns different from visual evoked cortical activity. That is, we expected that the MEG signals time-locked to the cardiac cycle onset would make the visual evoked effect disappear and vice versa.

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Results

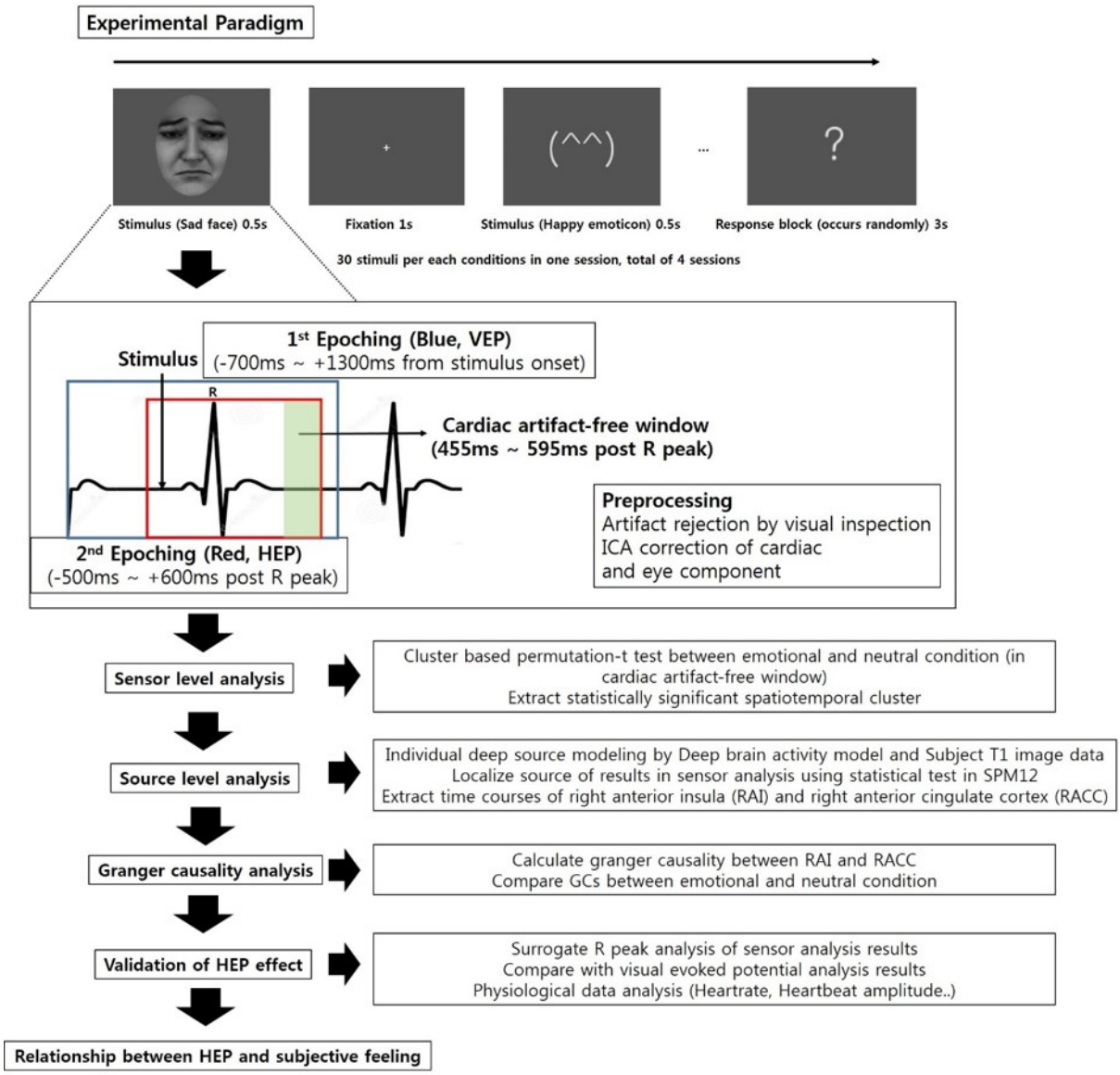

Fig. 1. Overall experimental flow. The sad face used in this figure is different from the one that used in the experiment (the sad face in this figure was generated by FaceGen Modeller (http://www.facegen.com/)).

Sensor analysis

Six conditions of stimuli consisted of happy face, sad face, neutral face, happy emoticon, sad

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

emoticon, neutral emoticon were presented to thirty-two participants while recording magnetoencephalography (MEG). Data were preprocessed including independent component analysis (Hyvärinen, Karhunen, & Oja, 2004) which could minimize cardiac field artifact (CFA). Then we compared HEP of emotional conditions and neutral conditions using cluster-based permutation paired *t* test (Oostenveld, Fries, Maris, & Schoffelen, 2011) in CFA-free time windows of 455ms to 595ms post R peak (Gray et al., 2007; Müller et al., 2015). Four tests were performed including a sad face vs. a neutral face, a happy face vs. a neutral face, a sad emoticon vs. a neutral emoticon, and a happy emoticon vs. a neutral emoticon.

Significant difference in the HEP between sad face and neutral face perception

A HEP cluster showing a significant difference between the perception of a sad face and that of a neutral face was found in the right frontocentral sensors within a 488ms-515ms time range (Monte-Carlo p = 0.046). In other conditions, including happy face vs neutral face (Monte-Carlo p = 0.396), sad emoticon vs neutral emoticon (Monte-Carlo p = 0.857), happy emoticon vs neutral emoticon (Monte-Carlo p = 0.710), there were no significantly different clusters showing a different HEP amplitude. Using the ROI-TOI mask of time and sensors of significant clusters found in the sad face vs neutral face condition, we performed an additional paired *t* test between the happy face or emotional emoticon condition and the neutral condition. In these analyses, only the happy face showed a significant difference (t (31) = -2.12, p = 0.042). Furthermore, we performed a correlation analysis with the HEP modulation effect of a happy face and a sad face (averaged within ROI-TOI mask) with PHQ-9 score (Supplementary materials). The HEP modulation was significantly correlated with participants' PHQ-9 score (Spearman's rho = 0.377, p = 0.033) in the happy face. Moreover, the HEP modulation effect of the happy face was more significant in participants with a low PHQ-9 score (t(18) = -3.314, p = 0.004) than in the paired t test within the whole group (t (31) = -2.12, p = 0.042), while

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

participants with a high PHQ-9 score did not show the HEP modulation effect ($t(12) = 0.30$, $p = 0.766$). The HEP modulation of the sad face was only marginally correlated with PHQ-9 (Spearman's rho = 0.297, $p = 0.099$). Detailed information of the analysis of the relationship between mood and HEP modulation is provided in the Supplementary materials.

Although the emotionality scores for a sad face and a sad emoticon were not different (in paired-t test, $p = 0.492$, mean emotionality score of sad emoticon = 2.81, mean emotionality score of sad face = 2.71), the sad emoticon did not show HEP modulation even in this analysis ($t (31) = -0.2506$, $p = 0.8038$). The happy emoticon also did not show HEP modulation ($t (31) = 0.048$, $p = 0.9724$).

Unlike face perception, which is an innate process, emoticon perception is an acquired process. Moreover, it has more abstract information than facial expression which needs an interpretation or appraisal process. Therefore, perception and interpretation of emotional emoticons are more likely to vary with a person's emotional state and emotional experience than perception of emotional faces. We performed an additional analysis controlling for individual differences in emotional experience like verbal abuse (detailed results and methods are provided in the Supplementary materials). Briefly, the HEP modulation effect of sad emoticons was found to be significantly affected by the degree of experience of peer verbal abuse (PeVA) in left central sensors (Monte-Carlo $p = 0.015$, $n = 30$). The HEP modulation effect from sad emoticons within this cluster appeared only in participants who had no history of peer verbal abuse (in paired t test using permutation method, Monte-Carlo $p = 0.004$, $n = 10$), while there was no significant HEP modulation in participants with verbal abuse history (Monte-Carlo $p = 0.212$, $n = 20$). Happy emoticons did not show any relationship with verbal abuse history (PeVA and PaVA) or mood (PHQ-9) (failed to form cluster in cluster based permutation regression analysis).

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

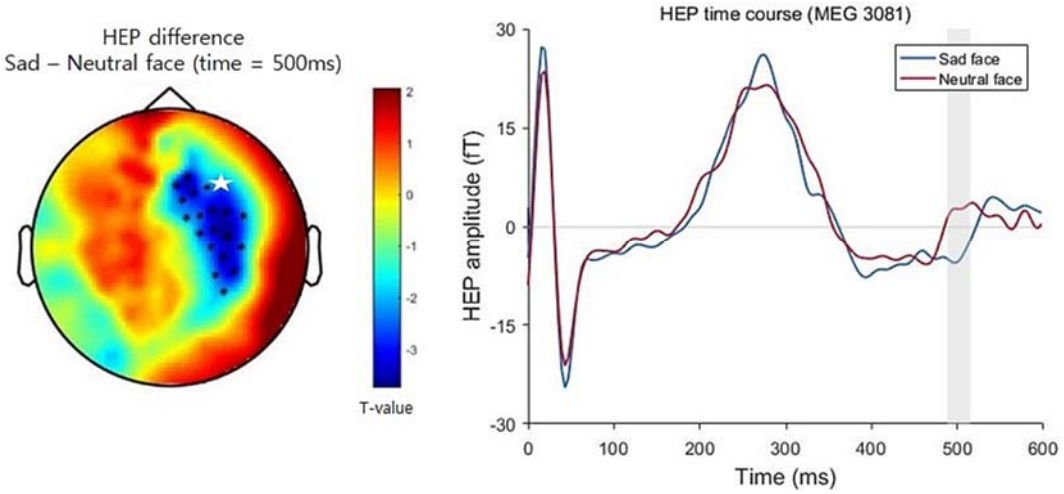

Fig. 2. Topographic map (left) of the differences in HEP between sad vs. neutral face conditions. Cluster showing significant difference is marked as black dot (Monte-Carlo p < 0.05, cluster corrected). Single channel plot among significant clusters (right), shaded area represents cluster of time point that shows significantly different time course between conditions. The channel plotted in the right figure are marked as a white star in the left topographic map.

Source analysis

Interoceptive network and prefrontal-basal ganglia network as sources of HEP modulation in sad faces

To find the brain region underlying modulation of the HEP in sad face conditions, source reconstructions of the MEG signal in a sad face and neutral face were done using the Brainstorm toolbox (Tadel, Baillet, Mosher, Pantazis, & Leahy, 2011). To estimate the time courses of both cortical and subcortical activity, we used the default settings in Brainstorm's implementation of the Deep Brain Activity model using the minimum norm estimate (MNE)

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(Attal & Schwartz, 2013; Tadel et al., 2011). Then, a paired *t* test of the two conditions was conducted using SPM12 (Penny, Friston, Ashburner, Kiebel, & Nichols, 2011)(with a spatial map that was averaged within the time range of 488ms-515ms after the R peak, which have shown significant different HEP between sad face and neutral face). With the cluster-forming threshold using p-value = 0.01 and 10 adjacent voxels, several regions that have different HEP time courses between a sad face and a neutral face were identified. Four significant clusters appeared. Briefly, these clusters included the right prefrontal cortices, the anterior insula, the anterior cingulate cortex, and the basal ganglia. More specifically, the first cluster (red in Fig. 3, $p = 0.003$, cluster level family-wise error (FWE) corrected) included the right prefrontal regions that consisted of the right superior frontal gyrus (RSFG), which is close to the dorsomedial prefrontal cortex (dmPFC), and the middle frontal gyrus (RMFG), which corresponded to the dorsolateral prefrontal cortex (dlPFC). The second cluster (green in Fig. 3, $p = 0.001$, cluster level FWE corrected) included the right anterior insula (RAI) and the right putamen (RP). The third cluster (blue in Fig. 3, $p = 0.003$, cluster level FWE corrected) included the right anterior cingulate cortex (RACC) and the left anterior cingulate cortex (LACC). Finally, the fourth cluster (pink in Fig. 3, $p < 0.001$, cluster level FWE corrected) included the right basal ganglia, which consisted of the right globus pallidus (RGP) and the right putamen (RP). Previous studies reported ACC and AI as a source of HEP (Couto et al., 2015; H.-D. Park et al., 2017; H.-D. Park, Correia, Ducorps, & Tallon-Baudry, 2014). Additionally, in a paired *t* test using the absolute value of time courses, right ventromedial prefrontal cortex (RvmPFC) and left orbitofrontal cortex (LOFC) clusters were found, which is also a region known to process interoceptive information; a detailed result of this analysis is provided in the Supplementary materials. Furthermore, we also performed the same analysis comparing a happy face and a neutral face, which showed the HEP modulation effect only in the paired *t* test using the ROI-TOI mask, to see whether the HEP modulation pattern induced

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

by a happy face was similar to that induced by a sad face. In the results, with the cluster-forming threshold using p-value = 0.01 and 10 adjacent voxels, only the right insula cluster was found to be different between the happy face and the neutral face, which was located in the mid-dorsal position of the insula (MNI coordinate = (33, 8, 8), $p < 0.001$, cluster level FWE corrected).

Table 1. Clusters showing significant different time courses between sad face and neutral face conditions.

Cluster	Region	MNI coordinate (x, y, z)	Cluster size (voxel number)	F value	Cluster P value (FWE corrected)
Cluster 1 (Red)	Rt. SFG	21, 20, 42	311	23.13	0.003
	Rt. MFG	33, 18, 36		12.86	
Cluster 2 (Green)	Rt. AI	27, 24, 6	379	20.02	0.001
	Rt. Putamen	31, 6, 8		12.08	
Cluster 3 (Blue)	Rt. ACC	9, 40, 10	313	17.32	0.003
	Lt. ACC	-3, 22, 28		14.68	
Cluster 4 (Pink)	Rt. putamen	27, -8, 0	412	16.59	<0.001
	Rt. GP	21, 0, 2		15.30	

SFG: superior frontal gyrus, AI: anterior insula, MFG: middle frontal gyrus, ACC: anterior cingulate cortex, GP: globus pallidus (voxel size: 1 mm x 1 mm x 1 mm)

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

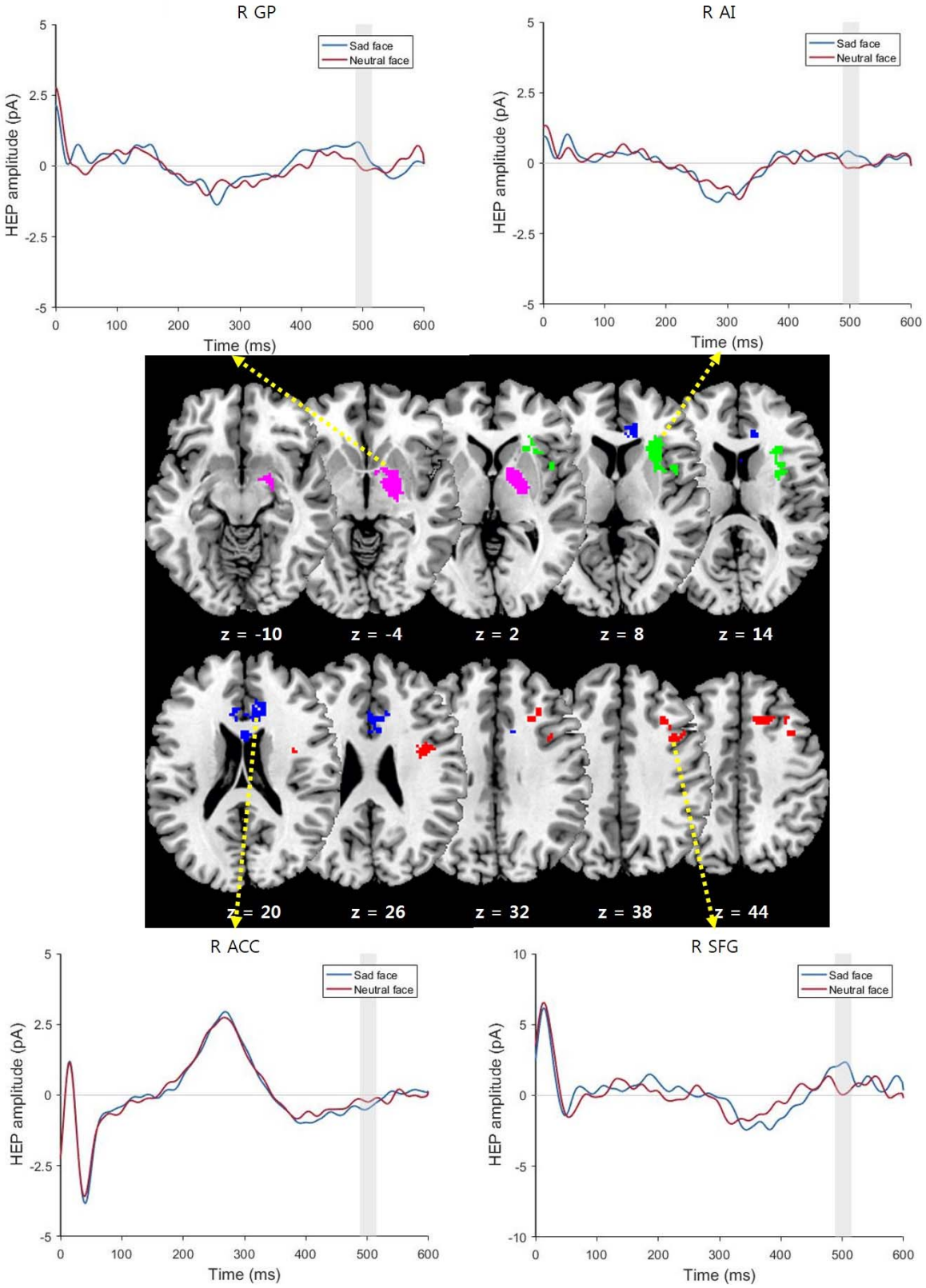

Fig. 3. Brain regions showing significant differences in HEP modulation in the contrast of sad vs. neutral face conditions ($p < 0.01$, cluster level family-wise error (FWE)

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

corrected) and the mean time courses of HEP from 25 voxels around peak voxels (the yellow dashed arrow connects the region of cluster and corresponding time courses of that region).

Granger causalities between sources of HEP modulation

The source analysis of HEP induced by a sad face revealed the activation of regions that were previously known to be key regions of feelings, namely, AI and ACC (Medford & Critchley, 2010; Smith & Lane, 2015). More specifically, they were hypothesized to act as the input (AI) and the output component (ACC) of a system in which the ACC re-represents interoceptive information relayed from the AI (Medford & Critchley, 2010). We hypothesized that when a person processes a sad face, causal information flow from the RAI to the RACC would be increased, while that of the RACC to the RAI remains at the baseline level. To test this hypothesis, we performed Granger causality (GC) analysis using MVGC toolbox (Barnett & Seth, 2014) on these two regions in both sad face and neutral face condition. Short time window GC estimation with a sliding window (Ding, Bressler, Yang, & Liang, 2000; Seth, Barrett, & Barnett, 2015) was used which is an appropriate method for the analysis whose temporal precision is important like a HEP modulation effect. Pairwise GC was calculated between the RAI and the RACC from the [455ms 515ms] window to the [505ms 565ms] window. Then we compared GCs of a sad face and a neutral face conditions for all time windows using cluster-based permutation paired-*t* test. Results showed that only the GC of HEP from the RAI to the RACC was significantly increased in the sad face compared to the neutral face from the 474ms to 568ms ([480ms 540ms] window to the [505ms 565ms] window, Monte-Carlo $p = 0.014$, cluster corrected for 2 GCs and 27 time windows). GC from the RACC to the RAI did not survive in multiple comparisons (Monte-Carlo $p = 0.146$, cluster from [493ms 553ms] to

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[501ms to 561ms]). These results indicate that only bottom up information from the RAI to the RACC was increased. Additionally, we analyzed GC change in a happy face between the right insula (found in the source analysis of the happy face) and RACC to determine whether a happy face has a similar GC change pattern as a sad face. An analysis of a happy face GC was performed in a way similar to the sad face, except that instead of the RAI, we used the peak coordinate of the right insula that corresponded to the mid-dorsal position of the insula, which was found in the source analysis of a happy face. The time course of the RACC was extracted from the same coordinate that was used in the sad face GC analysis because we assumed that even though there was no activation of the RACC in the happy face, there could be an information flow between those areas. Interestingly, the happy face had a slightly different pattern from the sad face, which means a top-down pattern of GC from the RACC to the right insula. GC from the RACC to the right insula were significantly increased compared to the neutral face from the [456ms 516ms] window to the [484ms 544ms] time window (Monte-Carlo $p = 0.013$) while there was no significant cluster time window in the right insula to the RACC (Monte-Carlo $p = 0.191$, cluster from [487ms 547ms] to [493ms to 553ms]).

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

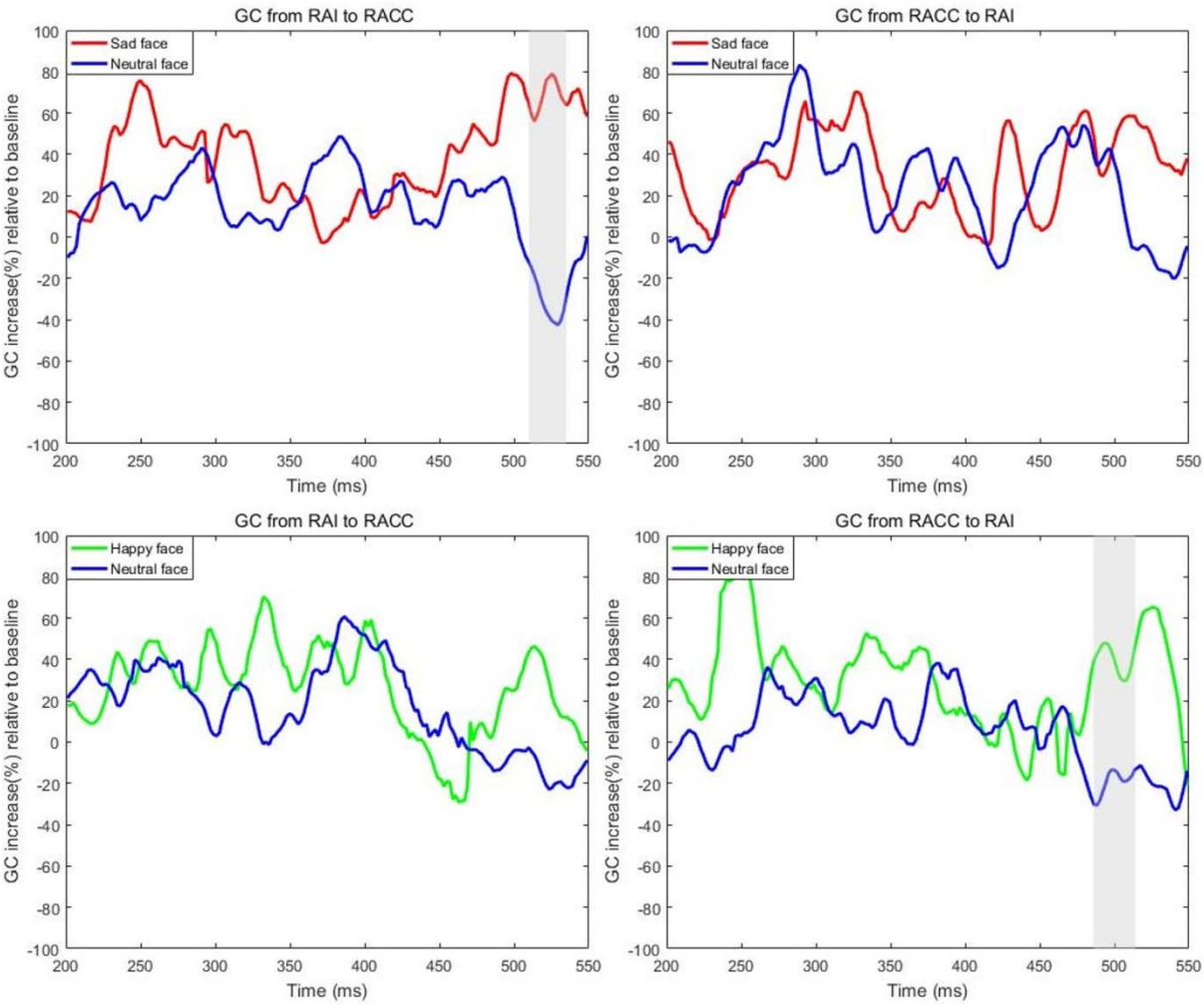

Fig. 4. GC analysis results. Increased bottom-up transfer of interoceptive information from the RAI to the RACC in sad face perception (left upper), and increased top-down information transfer from the RACC to the right insula (right lower). Time in figure represents the center time of sliding windows post R peak and the shaded area represents time windows that showed a significant difference between conditions after cluster correction (Monte-Carlo $p < 0.05$, cluster corrected). GCs from [170ms 230ms] window to [425ms485ms] window that was not included in the analysis are also plotted for visualization.

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Analysis of physiological data

Heart rate and ECG R peak amplitude modulation in each condition

To determine whether the effect of the different HEP amplitude originated from different heart-related physiological statuses, the heart rate and ECG R peak amplitude were compared between the conditions using one-way RANOVA. There were no significant differences between conditions in either heart rate (F (1.516, 46.982) = 2.367, p = 0.118, Greenhouse-Geisser corrected) or heartbeat amplitude (F (2.989, 92.658) = 0.958, p = 0.416, Greenhouse-Geisser corrected).

Heartbeat distribution in each condition

To rule out the possibility that our significant HEP modulation effect resulted from the fact that the heartbeat appeared more or less in a specific time window of visual evoked cortical processing, distribution of heartbeat occurrences within each time window of visual processing was analyzed. Two-way 6*9 RANOVA (six condition and nine 100ms time bins (-200ms from 700ms to visual stimulus onset)) showed neither a different occurrence rate of heartbeat among the conditions (F (3.810, 118.111) = 0.783, p = 0. 533, Greenhouse-Geisser corrected) nor among time bins (F (4.876, 151.143) = 1.540, p = 0.182, Greenhouse-Geisser corrected).

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

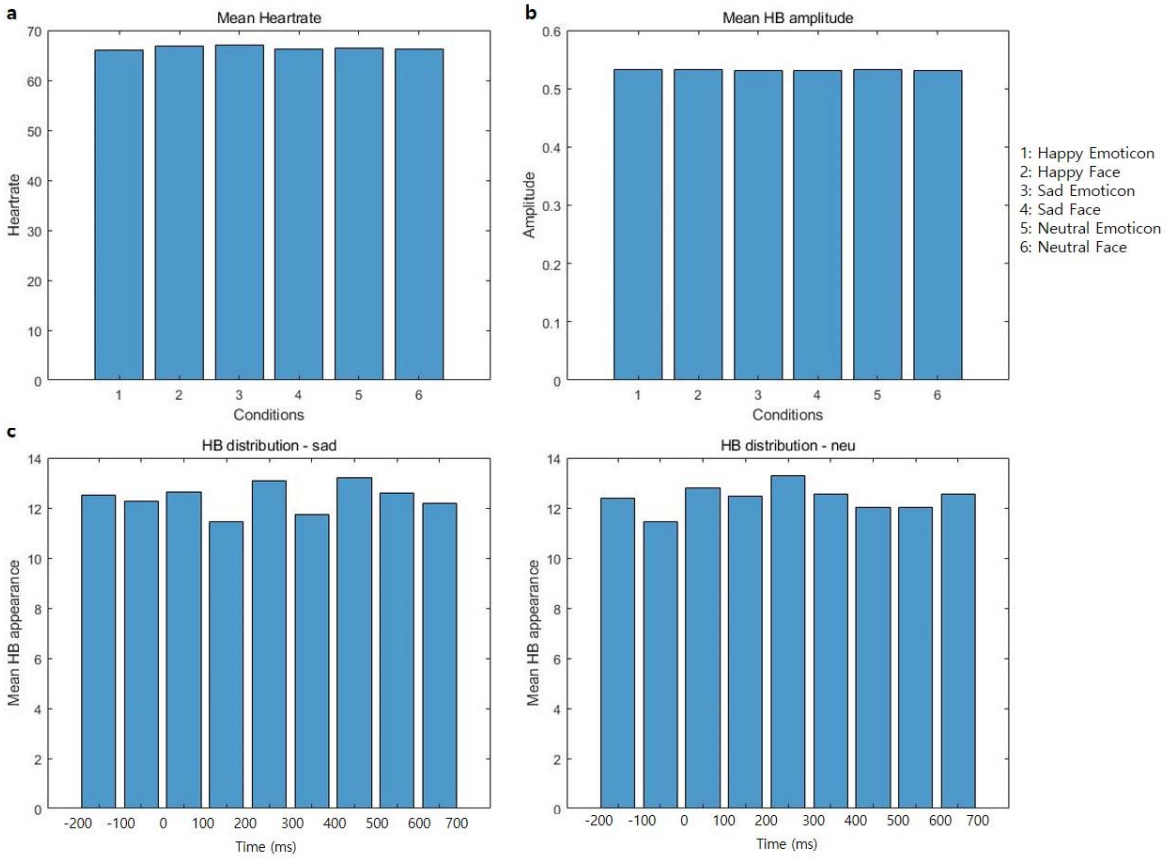

Fig. 5. Physiological data analysis results - a. mean heart rate of six experimental conditions b. mean R peak amplitude of six experimental conditions, c. mean R peak occurrence distribution in epoch of sad face (left) and neutral face (right).

Analysis to exclude the effect of visual processing

Surrogate R peak analysis on HEP modulation effect of a sad face

To test whether the HEP modulation effect is time-locked to the heartbeat, we created 100 surrogate heartbeats that were independent of original heartbeats (H.-D. Park et al., 2016; H.-D. Park et al., 2014). Then we computed the surrogate HEP with surrogate heartbeats and performed the same cluster-based permutation *t* test between conditions that showed a significant difference in the sensor level analysis. Then, the maximum summed cluster t

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

statistics of the difference between sad and neutral faces was calculated and used to make a permutation distribution. Our HEP modulation effect size (maximum cluster t statistics) was significant in that the distribution (Monte-Carlo p<0.03) indicating our effect was highly likely to be locked to the heartbeat.

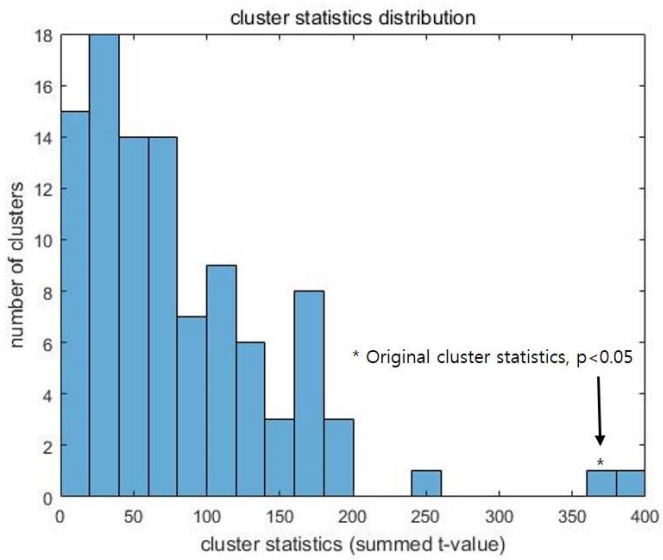

Fig. 6. The surrogate R peak histogram (absolute cluster statistics), * represents the location of the original cluster statistics of the HEP modulation effect in a sad face.

Different spatial pattern between visual evoked potential analysis (VEP) results and HEP analysis results in sad face perception

Finally, to exclude the possibility that HEP modulation was confounded by neural activity reflecting visual processing, we analyzed visual evoked potential data with a cluster-based permutation *t* test. Two significant clusters at 73ms-198ms (Monte-Carlo p = 0.05) and 815ms-948ms (Monte-Carlo p = 0.004) after stimulus were found. However, their topological distribution was totally different from the HEP effect. Furthermore, in the source analysis of VEP (which was performed with the same method as the HEP source analysis), two clusters, which included the right ventromedial prefrontal cortex cluster and right cuneus cluster, were

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

found at 73ms-198ms (cluster-forming threshold = 0.01, minimum number of adjacent significant voxels = 10, cluster level p-value < 0.001 in both clusters). Two clusters were also found in the 815ms-948ms time window, which included the left cuneus cluster and the right cuneus cluster ($p < 0.001$, $p = 0.042$ each with the same cluster definition of the earlier cluster). Finally, the paired *t* test of absolute time courses revealed that the right fusiform/middle temporal gyrus cluster (cluster level $p < 0.001$) and the right angular gyrus/superior temporal gyrus cluster (cluster level $p = 0.043$, FWE corrected) were more activated in the sad face of the 73ms-198ms window, while a supplementary motor area cluster (cluster level p-value = 0.003, FWE corrected) was found in the late time window (815ms-948ms). Detailed information on the absolute time courses, including statistical information, is provided in the Supplementary materials.

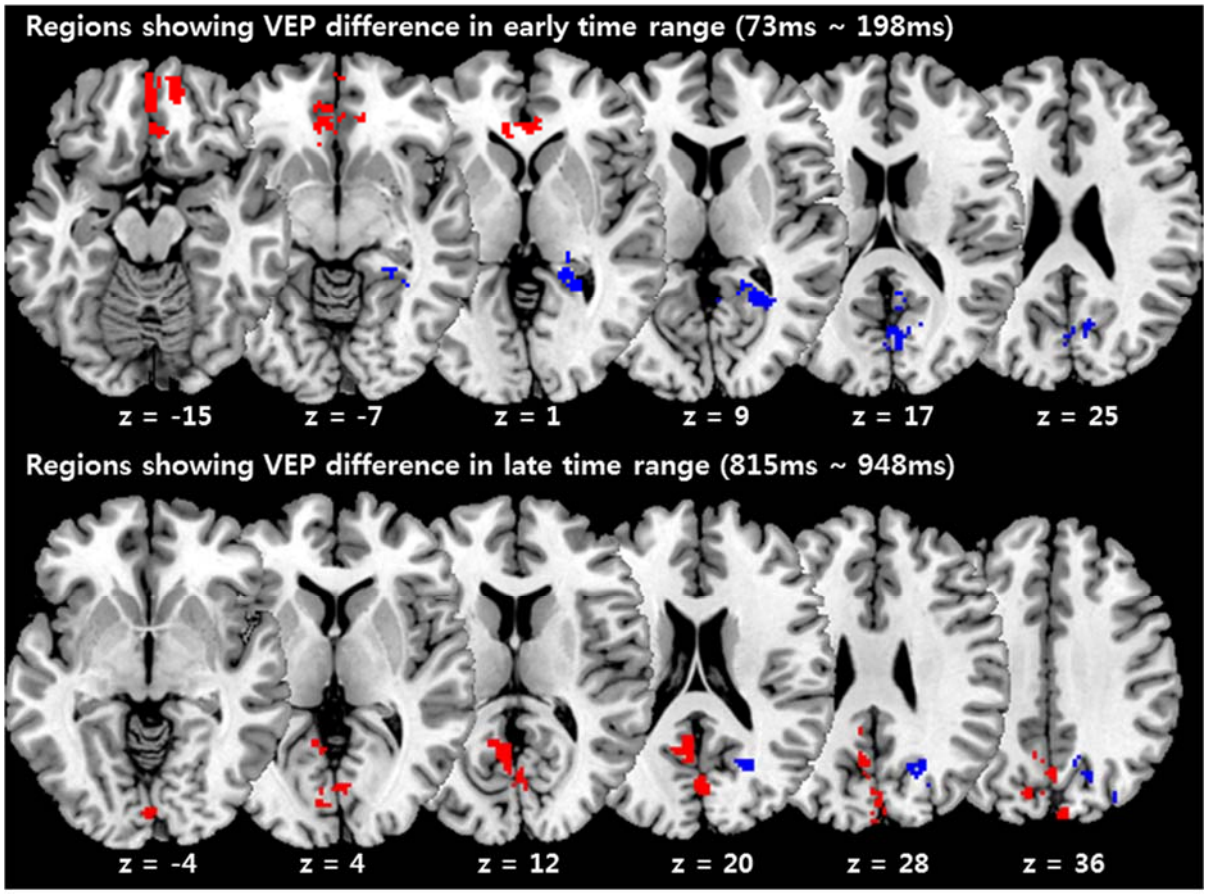

Fig. 7. Regions showing significantly different VEP for a sad face compared to a neutral

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

face ($p < 0.01$, cluster level family-wise error (FWE) corrected, upper: 73ms-198ms, lower: 815ms-948ms).

Table 2. Regions showing significantly different VEP for a sad face compared to a neutral face (73ms-198ms, 815ms-948ms)

Cluster (73ms-198ms)	**Region**	**MNI coordinate (x, y, z)**	**Cluster size (voxel number)**	**F value**	**Cluster P value (FWE corrected)**
Cluster 1 (Red)	Rt superior orbital gyrus	11, 56, -17	576	18.18	<0.001
	Rt medial orbital gyrus	11, 36, -10		15.59	
	Lt gyrus rectus	-1, 50, -15		14.65	
Cluster 2 (Blue) Rt Cuneus	Rt cuneus	3, -73, 18	465	18.17	<0.001
	Rt cuneus	11, -68, 26		14.48	
	Rt calcarine sulcus	23, -48, 6		14.18	
Cluster (815ms-948ms)	**Region**	**MNI coordinate**	**Cluster size (voxel number)**	**F value**	**Cluster P value (FWE corrected)**
Cluster 1 (Red)	Lt calcarine sulcus	-3, -62, 14	836	33.02	<0.001
	Lt cuneus	-1, -73, 20		25.06	

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

	Lt calcarine sulcus	-1, -87, -4		23.78	
Cluster 2 (Blue)	Rt cuneus	23, -68, 24	199	20.46	0.042
Rt Cuneus	Rt superior occipital gyrus	28, -78, 34		15.89	
	Rt superior occipital gyrus	23, -73, 30		12.15	

(voxel size: 1 mm x 1 mm x 1 mm)

Discussion

Our findings provide direct and strong evidence that perception of an emotional, especially sad, face modulates the interoceptive information processing in the cortex.

First, we showed that cortical heartbeat processing after presentation of a sad face has significantly different spatiotemporal dynamics compared with a neutral face and these differences are localized in the interoceptive network (AI, ACC, vmPFC), basal ganglia (GP, Putamen) and prefrontal areas (MFG, SFG). Importantly, results of the GC analysis of these regions showed that bottom up heartbeat information processing from the RAI to the RACC was increased in sad face condition. Moreover, although it might not be statistically rigorous, a happy face showed the HEP modulation effect in ROI-TOI mask analysis. Consistent with this result, the HEP modulation effect was localized in the right insula and the top down GC from the RACC to the right insula was increased in the GC analysis for happy face condition.

In contrast to the HEP results, visual – locked activity was different in the bilateral visual information processing area including the bilateral cuneus, fusiform face area and other areas

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

including the ventromedial prefrontal cortex and supplementary motor area. Interestingly, the only region that overlapped between the HEP and VEP were vmPFC – which is a key region in somatic marker hypothesis (Bechara et al., 2005). Finally, surrogate R peak analysis provided strong evidence that our result was a consequence of cortical heartbeat processing modulation. Additionally, analysis of physiological data and cardiac artifact removal using ICA also ruled out the possibility of an effect of other physiological effects on the cortical signal.

Our results go beyond previous studies of interoception and emotion in several aspects.

The results of our sensor analysis showed that a sad face evoked a different spatiotemporal pattern of HEP in the right frontal and central sensors at a time window centered about 500ms after the R peak, which was different from the pattern for a neutral face. Two recent studies with electroencephalography (EEG) reported HEP modulation by emotional stimuli. One study using visual and auditory stimuli showed the HEP modulation by high-arousal mood induction in left parietal clusters at 305 to 360ms after the R peak and the right temporoparietal cluster at 380 to 460ms after the R peak (Luft & Bhattacharya, 2015). They summed both positive and negative emotional valence conditions to show the arousal effect which is not specific to emotional feeling. Another recent EEG study in 5-month-old infants reported that a video clip of an angry or fearful face increased HEP at 150 to 300ms after R peak in the frontal cluster (Maister, Tang, & Tsakiris, 2017). While both happy and sad stimuli are low-arousal emotions (Luft & Bhattacharya, 2015), the second study showed neither a significant cluster in the contrast of happiness vs. neutral conditions nor applied sad stimuli. HEP modulations in these two studies more likely reflected the representation of body signals induced by arousal than the conceptualizing process of specific emotions. In addition, neither of these studies demonstrated the underlying neural mechanism of this modulation effect based on source analysis. Considering that the bodily signals not only need to be recognized in the brain but

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

also to be conceptualized to make emotional feeling (Smith & Lane, 2015), we contend that our results of HEP modulation by a sad face especially increasing GC from the RAI to the RACC reflects the conceptualization process of interoceptive signals that are related to specific emotional feelings, not just a representation of bodily arousal signals. First, the ACC is thought to be related to forming an emotional concept based on whole body representation of the anterior insula (Medford & Critchley, 2010; Smith & Lane, 2015), and representation of body signals in the anterior insula is known to be formed by processing of the primary interoceptive representation processed by the posterior insula (Craig, 2009; Smith & Lane, 2015). In our main result, HEP modulation effect showing in the RAI and the RACC was not found in the RPI. Furthermore, additional pairwise GC analysis on three regions of interest, the RPI, RAI, and RACC, confirmed increased GC from the RAI to the RACC, but failed to show any modulation of GC from RPI to RAI by a sad face (details of the analysis is described in the Supplementary materials). These suggest that the HEP modulation effect in a sad face might not be related to primary interoceptive processing of physiological arousal; rather, it is likely to be related to later processing starting from the anterior insula to the anterior cingulate. Moreover, the spatiotemporal dynamics of our results are later than HEP modulation by both a heartbeat perception task, which typically appears 250ms ~ 250ms after the R peak (Pollatos & Schandry, 2004), and by arousal that appeared before 460ms post R peak (Luft & Bhattacharya, 2015). Finally, in one fMRI study of inducing sad feeling status showed an increased functional connectivity between the RAI and the RACC, which also supports our results on HEP modulation by sad face (B. J. Harrison et al., 2008).

Considering these aspects, our results are not likely to be caused only by representation of primary arousal signals and are likely to be related to later processing, which we suggest is related to conceptualization and creation of individual emotional feeling.

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Contrary to a sad face, a happy face showed HEP modulation only in the ROI-TOI mask, which seems much weaker than HEP modulation of a sad face and it is hard to conclude whether HEP modulation existed. However, considering the results of the source analysis that localized the HEP modulation effect in the right mid-dorsal portion of the insula and results of GC analysis, which showed an increased GC from the RACC to the right insula, we suggest that there is a HEP modulation effect for a happy face, while the pattern of modulation is slightly different from that of a sad face. This result is also supported by the previous fMRI study of happy facial expression that showed activation of location similar to that in our study (Pohl, Anders, Schulte-Rüther, Mathiak, & Kircher, 2013). Furthermore, previous theories hypothesized that depression is related to interoceptive prediction (Paulus & Stein, 2010) and also HEP was differ between depressed and nondepressed groups while performing heartbeat perception tasks (Terhaar et al., 2012). In our results, HEP modulation of happy faces seemed to be influenced largely by a person's mood state, which was reflected in a significant correlation between HEP modulation in the ROI-TOI mask and the PHQ-9 score, while HEP modulation of a sad face had only a marginal correlation. Considering previous studies and our results, it seems that a depressive mood largely affects interoceptive processing related to happy faces, which might have caused the weak HEP modulation effect in our results.

Our results showing both a significant relationship of the HEP difference between a sad emoticon and a neutral condition with the peer verbal abuse score across all subjects and a significant difference of the HEP difference in participants with no peer verbal abuse history suggest that the perception of a sad emoticon also modulates the interoceptive information processing that is influenced by subjective emotional experience such as peer verbal abuse. Although this might not be a complete explanation for the lack of the HEP modulation effect with a sad emoticon in contrast to a sad face, considering that there are not many studies related to the effect of emotional experience on responses to emoticons, our results might provide some

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

clue to the effect of a person's emotional experience on their emotional response to acquired (not innate) emotional stimuli. Further studies are needed. Finally, happy emoticons have shown neither the HEP modulation effect nor a relationship to psychological scores, like verbal abuse or mood. However, it might be too hasty to conclude that a happy emoticon does not modulate HEP considering that we did not control for positive emotional experience, which might influence perception of a happy emoticon, in a way similar to the influence of negative emotional experience on sad emoticon perception. Future study investigating influence of positive emotional experience on happy emoticon perception would clarify this point.

In source analysis of HEP and VEP modulation by a sad face, we found that there are two clearly distinct systems of processing, cardiac information processing and visual information processing. First, brain regions that reflect HEP modulation in sensor analysis were found in the RAI and the RACC, which is a previously known source of HEP (Couto et al., 2015; H.-D. Park et al., 2017; H.-D. Park et al., 2014; Pollatos, Kirsch, & Schandry, 2005). These regions are also identified as overlapping regions of emotion, interoception and social cognition in a recent meta-analysis (Adolfi et al., 2016). Moreover, a Granger causality analysis revealed that cardiac information flow from the RAI to the RACC increases with sad face perception. The anterior insula interacts with many regions, including the ACC, which is a central autonomic network that regulates autonomic function (Beissner, Meissner, Bär, & Napadow, 2013). Furthermore, the RAI to the RACC is proven to have a causal relationship in resting state fMRI in GC analysis, while the reverse directional connectivity is weak (Sridharan, Levitin, & Menon, 2008). Our results suggest that this causality might be increased in sad face perception.

These previous studies make it clear that bottom-up cardiac information flow is increased in sad face perception. However, according to the source analysis of VEP, only the visual cortical region and ventromedial prefrontal regions appeared, which is consistent with previous

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

EEG/MEG studies of sad faces (Batty & Taylor, 2003; Esslen, Pascual-Marqui, Hell, Kochi, & Lehmann, 2004). By integrating these results, we firmly insist that processing of sad faces involves distinct interoceptive processing and visual processing and this is revealed by the HEP and VEP after the stimulus. To our knowledge, this is the first study that has shown distinctive processing of both systems of processing induced by sad faces in a clear and direct way. These results correspond well to hypotheses that explain the relationship between emotions and interoception, like the somatic marker hypothesis (A. R. Damasio, 2003), which predicts that bottom up interoceptive processing is increased by emotional stimulus. The somatic marker hypothesis predicts that physiological change is induced by emotional stimulus and modulates brain-body interaction, while the results of our physiological data analysis showed that cardiac activity parameters, including heart rate and heartbeat amplitude, were not modulated by the sad face. However, considering that direct input of the HEP is pressure in the carotid baroreceptor and that we did not measure blood pressure or index of carotid stimulation, it is hard to tell whether there was physiological change. Another possibility is that absence of induced physiological change might be due to our short stimulus onset asynchrony of about 1 second, in which it is hard to evaluate physiological changes like heart rate modulation, which occurs after several seconds (Critchley et al., 2005). Note that our results were derived from a sad emotion among negative ones. Therefore, future experiments need to be performed with other negative emotional stimuli, such as fear or anger.

An unexpected finding in the present study was that cardiac information processing was also modulated in the basal ganglia (RGP/RP) and prefrontal regions (SFG/MFG). Globus pallidus is known to send input to the prefrontal cortex via the thalamus. This pathway is related to initiating motor action (Singh-Bains, Waldvogel, & Faull, 2016). In particular, the ventral pallidum (VP) is closely related to regulating emotion or starting motor action in response to emotional stimuli (Singh-Bains et al., 2016). Moreover, there was a case of a patient with

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

damage of the GP, including the VP, who reported inability to feel emotion (Vijayaraghavan, Vaidya, Humphreys, Beglinger, & Paradiso, 2008). Based on this evidence, we suggest that cardiac information is relayed to the GP (including the VP), and finally, the prefrontal area has a role in initiating emotion-related behavior like facial expression or generation of feeling triggered by cardiac information processing, which is consistent with the somatic marker hypothesis.

To our knowledge, this is the first study to show different interoceptive and visual processing by the same emotional stimulus. In conclusion, our results demonstrate that processing of sad faces induces different interoceptive information processing and visual processing compared to neutral faces, which is reflected by the HEP and VEP, respectively. Interoceptive processing involves increased bottom-up processing of the HEP from the RAI to the RACC, while different visual processing occurs in a different area, including the visual cortical area. Additionally, we found that cardiac signals are also processed differently in the basal ganglia and prefrontal regions in sad face processing, which might reflect the initiation of emotion-related behavior like facial expression or generation of feeling.

Methods

Participants

Forty healthy participants (19 females, mean age of 24.03 ± 3.28 years) volunteered in the experiment. The expected effect sizes were not known in advance, so we chose a sample size of approximately 40 participants, which was approximately two times more than those of previous MEG and EEG studies of HEP (Babo-Rebelo, Richter, & Tallon-Baudry, 2016; Fukushima et al., 2011; H.-D. Park et al., 2014).

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

MEG recording consisting of 4 runs was completed with one visit. High resolution T1 weighted MRI scans were acquired at another visit. In this MRI session, all subjects performed both functional MRI experiments consisting of emotion discrimination (unpublished) and/or decision (unpublished) tasks and other structural MRIs such as diffusion tensor imaging (unpublished). Among the forty subjects, we failed to acquire MEG data for five subjects due to magnetic field instability. Another three subjects were excluded in analysis because their ECG data were too noisy or absent. Therefore, thirty-two subjects were included for further analysis.

A structured interview was conducted using the Korean version of the diagnostic interview for genetic studies (Joo et al., 2004). None of the subjects had current neurological or psychological disease. Subjects completed Patient Health Questionnaire-9 (PHQ-9)(Kroenke, Spitzer, & Williams, 2001), a self-rating scale for depression, and the Korean version of the Verbal Abuse Questionnaire (K-VAQ) (Jeong et al., 2015) because we suspected that previous emotional abuse history might affect response to emotional stimulus. A mean of PHQ-9 score was 3.62 (SD: 2.59, n=39). The mean score of peers (PeVA) and parents (PaVA) were 22.61 (SD: 9.65, n=38) and 21.95 (SD: 7.82, n=38), respectively. All thirty-two participants who were included in the analysis completed PHQ-9 (mean and SD: 3.68 and 2.69, n=32), while two participants among them did not complete K-VAQ (mean and SD of PeVA: 21.73 and 7.26, mean and SD of PaVA: 23.43 and 10.29, n = 30). All participants submitted written informed consent to participate in the experiment. The study was approved by the Korean Advanced Institute of Science and Technology Institutional Review Boards in accordance with the Declaration of Helsinki.

Standardization of emotional stimuli

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Stimuli consisted of forty-five emotional faces and forty-five text-based emotional emoticons. Forty-five faces expressing happy, sad, and neutral emotions were selected from the Korean Facial Expressions of Emoticon (KOFEE) database (J. Y. Park et al., 2011). Text-based happy and sad emoticons were searched for on the world-wide web. Then we produced scrambled emoticons that did not have configurable information and used these scrambled emoticons as emoticons of a neutral condition. Ninety emotional expressions, including faces and text-based emoticons, were standardized in independent samples consisting of forty-seven healthy volunteers (21 females, mean age of 28.43 ± 4.31 years). These participants were asked to rate the feeling they felt toward emotional expressions composed of 90 stimuli (45 faces and 45 facial emoticons of happy, neutral, and sad emotions) on an 11 point Likert scale (-5 to +5). We compared the mean absolute value of the four emotional expressions that we called 'feeling intensity' or 'emotionality' (Citron, Gray, Critchley, Weekes, & Ferstl, 2014). Repeated measures analysis of variance (RANOVA) of 2 (face, emoticon) by 3 valences (happy, sad, neutral) design was performed on the mean and variance of emotionality score, respectively. In the RANOVA of mean, there was a significant main effect of valence ($F(1.744, 80.228) = 272.618$, $P < 0.001$, Greenhouse-Geisser corrected), while there were no differences between emoticons and faces ($F(1, 46) = 0.011$, $P = 0.919$) and no interaction between those two main effects ($F(1.685, 77.488) = 0.285$, $P = 0.818$, Greenhouse-Geisser corrected). In addition, a post-hoc *t* test revealed that there was no difference between sad and happy conditions ($P = 0.082$) with a significant difference between emotional and neutral conditions ($P < 0.001$ in both sad and happy contrast to neutral). In the RANOVA of variance, the variance of emoticons was significantly larger than the variance of faces ($F(1, 46) = 16.108$, $P < 0.001$), while there were no difference in the variance between emotions ($F(1.268, 58.342) = 2.608$, $P = 0.079$, Greenhouse-Geisser corrected) and no significant interaction ($F(1.347, 61.963) = 4.831$, $P = 0.066$, Greenhouse-Geisser corrected). Additionally, participants of the main experiments also

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

underwent the rating procedure above before the MEG recording.

MEG experimental task

During the MEG recording, ninety stimuli consisting of 45 faces and 45 text-based emoticons were presented in the center of the screen using in-house software, the KRISSMEG Stimulator 4. The size, duration, and stimulus onset asynchrony (SOA) of all the stimuli were 27×18 cm, 500ms and 1500ms, respectively, and the order of stimuli presentation was pseudo-randomized. Participants completed 4 runs and each run contained 180 stimuli (30 sad faces, 30 happy faces, 30 neutral faces, 30 sad emoticons, 30 happy emoticons, 30 neutral emoticons each) and took 270 s. In addition, to maintain the participants' attention to task, the participants were instructed to discriminate between sad and happy by pressing a button when a question mark appeared. The question mark randomly appeared on the screen every 9 to 15 trials.

Acquisition

A 152-channel MEG system (KRISS MEG, Daejeon, Korea, 152 axial first-order double-relaxation oscillation superconducting quantum interference device (DROS) gradiometers) covering the whole head was used to make MEG recordings in a magnetically shielded room for 60–90 minutes at a sampling rate of 1,024 Hz. The relative positions of the head and the MEG sensors were determined by attaching four small positioning coils to the head. The positions of the coils were recorded at intervals of 10–15 min by the MEG sensors to allow co-registration with individual anatomical MRI data. The maximum difference between head positions before and after the run was deviation < 2 mm and goodness of fit (GoF) > 95%. The EEG system for recordings of eye and muscle artifacts were made simultaneously with the

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

MEG recordings. During MEG recording, participants were seated with their heads leaning backward in the MEG helmet. The translation between the MEG coordinate systems and each participant's structural MRI was made using four head position coils placed on the scalp and fiducial landmarks (Hämäläinen, Hari, Ilmoniemi, Knuutila, & Lounasmaa, 1993).

Data preprocessing

Data were processed with a Fieldtrip toolbox (Oostenveld et al., 2011). First, raw data were epoched from 700ms before stimulus onset to 1300ms after stimulus onset. Epochs containing large artifacts were rejected by visual inspection. After artifact trial rejection, the eye movement artifact and the cardiac field artifact were removed by ICA (Hyvärinen et al., 2004) using the function "ft_componentanalysis" with the runICA algorithm. Twenty components were identified for each of six conditions. Two neurologists visually inspected each component, and components that showed typical spatial and temporal patterns of the CFA, eye blinking and movement noise were removed. After removing the noise, the data were filtered with a 1-40 Hz Butterworth filter. Then, the heartbeat evoked potential (HEP) for each stimulus condition was extracted by subsequent epoching, which was time-locked to the R peak of every epoch. R peaks were detected using the Pan-Tompkins algorithm (Pan & Tompkins, 1985) and the HEP of each condition was extracted by epoching 500ms before the R peak to 600ms after the R peak in the epoch of each condition. Because a heartbeat enters the central nervous system (CNS) around 200ms after the R peak by vagal afferent stimulation at the carotid body (Eckberg & Sleight, 1992) and a visual stimulus enters the CNS immediately through the retina, a heartbeat that occurs before a -200ms visual stimulus onset stimulates the brain earlier than a visual stimulus onset. Therefore, we excluded R peaks that occurred before 200ms of a stimulus onset to include heartbeat evoked processing that occurred only after visual stimulus.

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Therefore, this procedure excluded cortical input of a heartbeat that occurred before the visual stimulus. The R peak after 700ms of stimulus onset was also excluded because that HEP epoch would contain the next visual stimulus onset. Finally, a baseline correction was performed using a pre-R-peak interval of 300ms and trials of the same condition for each subject were averaged. We expected that HEP modulation induced by emotional faces and emoticons would be much smaller than CFA. Therefore, to control CFA maximally, in addition to removal of CFA related IC in ICA, we only used a time window of 455-595ms after each R peak, which is known to have the minimal influence on the CFA (less than 1%) as in the previous study (Gray et al., 2007; Müller et al., 2015).

Sensor analysis: Cluster-based permutation paired *t* test between each emotional condition and neutral condition

We compared the HEP of an emotional condition and a neutral condition. Four tests were performed including a sad face vs. a neutral face, a happy face vs. a neutral face, a sad emoticon vs. a neutral emoticon, and a happy emoticon vs. a neutral emoticon. To deal with multiple comparison problems, we used a cluster-based permutation paired *t* test. These tests were done as follows. First, data were downsampled to 256 Hz to make the computation efficient and paired *t* tests were performed at all time points between 455 and 595ms and all sensors. Then significant spatiotemporal points of uncorrected p-values below 0.05 (two-tailed) were clustered by the spatiotemporal distance and the summed t-value of each cluster was calculated. After calculating the cluster t-stat, permutation distribution was made by switching condition labels within subjects randomly, calculating the t-value of the paired *t* test between permutated conditions, forming clusters as mentioned above, selecting the maximum cluster t-value and repeating this procedure 5000 times. Finally, after the maximum cluster t-values of each

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

permutation made the permutation distribution, the corrected p-value original clusters were calculated. Additional paired *t* tests were performed between conditions that did not show a significant difference in the cluster-based permutation *t* test (i.e., happy face vs neutral face, sad/happy emoticon vs neutral emoticon) using the mean HEP values within the significant clusters observed by the cluster-based permutation *t* test (i.e., sad face vs neutral face condition; ROI-TOI mask).

Source analysis

Source reconstruction was conducted using the MATLAB package Brainstorm (Tadel et al., 2011). To estimate the time courses of both cortical and subcortical activity, we used the default settings in open source Matlab toolbox Brainstorm's implementation of the Deep Brain Activity model using the minimum norm estimate (MNE) (Attal & Schwartz, 2013; Tadel et al., 2011). First, cortical surfaces and subcortical structures, including the amygdala and basal ganglia, were generated for each subject from 3T MPRAGE T1 images using Freesurfer (Fischl, 2012). The individual heads/parcellations were then read into Brainstorm (Tadel et al., 2011) along with track head points to refine the MRI registration. In Brainstorm, a mixed surface/volume model was generated, and 15,000 dipoles were generated on the cortical surface and another 15,000 dipoles were generated in the subcortical structure volume. Refining the registration with the head points improves the initial MRI/MEG registration by fitting the head points digitized at the MEG acquisition and the scalp surface. Using the individual T1 images and transformation matrix generated as above, a forward model was computed for each subject using a realistic overlapping spheres model. The source activity for each subject was computed using the MNE (Brainstorm default parameters were used). The source map was averaged over a time window of 488ms to 515ms - which showed a significant difference between a sad face

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

and a neutral face at a sensor level (other emotional conditions were not significantly different from neutral conditions in cluster based permutation paired *t* test). Then this averaged spatial map was exported to SPM12 software (Penny et al., 2011) and subsequent statistical tests were performed. Paired *t* tests were used to identify regions that had a different HEP time course within the selected time window between the sad face and the neutral face. Moreover, to test the absolute activation difference between the two conditions, we additionally exported the absolute spatial map to SPM12 and applied the paired *t* test design. In source analysis, data were downsampled to 512 Hz using a sufficient number of time points in the Granger causality analysis. Finally, we performed a paired *t* test between happy face and neutral face in the source that showed a significant difference in the additional paired *t* test of the sensor analysis.

Granger causality analysis of HEP source activity

After identifying brain regions that had different time courses, we performed a GC analysis (Barnett & Seth, 2014) on two regions of interest, the right anterior insula (RAI) and the right anterior cingulate cortex (RACC), to determine whether effective connectivity between these regions are modulated differently in emotional expression compared to the neutral condition. These regions are known to be core regions of interoceptive processing and feeling according to many previous studies (Adolfi et al., 2016; A. R. Damasio, 2003; Smith & Lane, 2015). Moreover, emotion processing in AI and ACC are known to be right lateralized, especially in AI (Craig, 2009; Gu, Hof, Friston, & Fan, 2013). Time courses of these two regions were extracted from voxels that showed a significant difference in source analysis. Detailed information including coordinates of ROIs are provided in Supplementary materials.

In GC (Granger, 1988) analysis, time course Y causes time course X if the past time points of Y and X explains X better than the past of X alone. It is formulated by the log-likelihood ratio

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

between the residual covariance matrix of the model that explains X by the past of X and Y and the residual covariance matrix of the model that explains X by the past of X alone (Barnett & Seth, 2014).

$$X_t = \sum_{k=1}^{p} A_{xx,k} * X_{t-k} + \sum_{k=1}^{p} A_{xy,k} * Y_{t-k} + \varepsilon_{x,t}$$

$$X_t = \sum_{k=1}^{p} A'_{xx,k} * X_{t-k} + \varepsilon'_{x,t}$$

$$F_{Y \to X} = \ln |\Sigma'_{xx}| / |\Sigma_{xx}|$$

A is the matrix of the regression coefficient, Epsilon is the residual, Sigma is the covariance matrix of the residual, and F is the GC of X and Y. All these calculations were done using a multivariate Granger causality toolbox (MVGC toolbox) (Barnett & Seth, 2014).

Time courses of two ROIs were extracted for every trial for each subject. To satisfy stationarity assumptions of GC analysis, we used short time window GC estimation with a sliding window (Ding et al., 2000; Seth et al., 2015). This approach was also appropriate for our analysis of the HEP modulation effect, which was found in quite a short time window, and thus needed high temporal precision of the GC estimation. The size of the window was 60ms and step size was 2ms (Cohen, 2014; Ding et al., 2000) and calculation of GC was done for whole epoch, which started from the [-500ms -440ms] window to the [540ms 600ms] window. Further control for stationarity was done by removing the average event-related potential from each trial (Wang, Chen, & Ding, 2008). Model order was determined using Akaike Information Criterion (AIC) to a maximum order of seven time points, which corresponded to 14ms. After model estimation, we tested the stationarity of the model by examining whether the spectral radius $\rho(A) > 1$ in every time window and every subject(Barnett & Seth, 2014). Although we tried to control every time window to satisfy the stationarity assumption, after the [507ms 567ms] time

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

window, 21 participants violated the stationarity assumption. Additionally, violation occurred at around 0ms post R peak in three participants and at 370ms (around the T peak) in one participant. There was a similar pattern even when we tested the variable length of the time window and model order. We suspected that this stationarity violation might be induced by CFA. Therefore, we again used the time window starting from the [455ms 515ms] window to the [507ms 567ms] window, and every time window satisfied the stationarity assumption in every participant. The pairwise GC of two ROIs (two GCs - RAI to RACC, RACC to RAI) were each calculated for emotional and neutral conditions. To compare between emotional and neutral conditions, GC baseline normalization was performed in both conditions by calculating the change in GC relative to the average GC between the [-330ms -270ms] window and the [-130ms -70ms] window (Cohen, 2014). Time windows around 0ms post R peak were not used as a baseline because three subjects violated the stationarity assumption. Finally, the 2 estimated GCs of emotional and neutral conditions were compared using a cluster based permutation paired *t* test for all time windows starting from the [455ms 515ms] window to the [507ms 567ms] window (Oostenveld et al., 2011). Therefore, the multiple comparison was controlled for the number of GCs and the number of time windows using a cluster-based permutation paired *t* test.

Analysis of physiological data

Heartbeat distribution in each condition

To show that the HEP effect is not the result of a biased heartbeat distribution, within the original visual epoch, we divided the visual epoch between -200ms and 700ms, which was the beginning and end of the HEP epoching, by 100ms time windows (a total of nine time bins) and counted how many heartbeats there were in those time windows. We did this for every

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

condition. Then, the analysis of variance between nine time bins was performed to test whether the occurrence of heartbeats was the same in every time bin.

Heart rate modulation in each condition

The heart rate in every condition was calculated for each subject. Then, an analysis of variance between every condition was performed to test whether the heart rates were different across conditions.

Analysis to exclude the effect of visual processing

Surrogate R peak analysis

To test whether the HEP modulation effect is time-locked to the heartbeat, we created 100 surrogate heartbeats that were independent of original heartbeats (H.-D. Park et al., 2016; H.-D. Park et al., 2014). Then we computed the surrogate HEP with surrogate heartbeats and performed the same cluster-based permutation *t* test between conditions that showed a significant difference in the sensor level analysis. Finally, we made the distribution of maximum cluster statistics of a surrogate R peak and the calculated position of our original cluster statistics in this distribution to show that the heartbeat-locked effect is significantly large in such a distribution.

Analysis of visual evoked potential (VEP)

To test whether the HEP modulation effect is confounded by the visual evoked potential effect, we performed the same cluster-based permutation test with a visual stimulus-locked potential and compared the topology of significant clusters between HEP clusters and the VEP sensor level. Then, we performed a source localization of the VEP activity in a significant cluster time window and exported it to SPM12 to perform a statistical test between emotional and neutral

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

conditions with the same methods we used in the HEP analysis (including the absolute value difference). Then we compared the resulting source with the result of the HEP analysis. This VEP analysis was not done for all conditions but for conditions that had a significant HEP modulation effect so that we could compare significant modulation of the HEP with the VEP effect.

Acknowledgement

This research was supported by the Brain Research Program through the National Research F oundation of Korea (NRF) funded by the Ministry of Science & ICT (NRF - 2016M3C7A1914448 NRF - 2017M3C7A1031331). The authors wish to acknowledge Kyung -Min An and Yong-Ho Lee for helping data acquisition.

References

Adolfi, F., Couto, B., Richter, F., Decety, J., Lopez, J., Sigman, M., . . . Ibáñez, A. (2016). Convergence of interoception, emotion, and social cognition: A twofold fMRI meta-analysis and lesion approach. *Cortex*.

Attal, Y., & Schwartz, D. (2013). Assessment of subcortical source localization using deep brain activity imaging model with minimum norm operators: a MEG study. *PLoS One, 8*(3), e59856.

Babo-Rebelo, M., Richter, C. G., & Tallon-Baudry, C. (2016). Neural responses to heartbeats in the default network encode the self in spontaneous thoughts. *Journal of Neuroscience, 36*(30), 7829-7840.

Barnett, L., & Seth, A. K. (2014). The MVGC multivariate Granger causality toolbox: a new approach to Granger-causal inference. *Journal of neuroscience methods, 223*, 50-68.

Batty, M., & Taylor, M. J. (2003). Early processing of the six basic facial emotional expressions. *Cognitive Brain Research, 17*(3), 613-620.

Beissner, F., Meissner, K., Bär, K.-J., & Napadow, V. (2013). The autonomic brain: an activation likelihood estimation meta-analysis for central processing of autonomic function. *Journal of Neuroscience, 33*(25), 10503-10511.

Citron, F. M., Gray, M. A., Critchley, H. D., Weekes, B. S., & Ferstl, E. C. (2014). Emotional valence and

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

arousal affect reading in an interactive way: neuroimaging evidence for an approach-withdrawal framework. *Neuropsychologia, 56*, 79-89.

Cohen, M. X. (2014). *Analyzing neural time series data: theory and practice*: MIT Press.

Couto, B., Adolfi, F., Velasquez, M., Mesow, M., Feinstein, J., Canales-Johnson, A., . . . Sigman, M. (2015). Heart evoked potential triggers brain responses to natural affective scenes: a preliminary study. *Autonomic Neuroscience, 193*, 132-137.

Craig, A. D. (2009). How do you feel—now? the anterior insula and human awareness. *Nature reviews neuroscience, 10*(1).

Critchley, H. D., Rotshtein, P., Nagai, Y., O'doherty, J., Mathias, C. J., & Dolan, R. J. (2005). Activity in the human brain predicting differential heart rate responses to emotional facial expressions. *Neuroimage, 24*(3), 751-762.

Damasio, A., & Carvalho, G. B. (2013). The nature of feelings: evolutionary and neurobiological origins. *Nature reviews neuroscience, 14*(2), 143-152.

Damasio, A. R. (2003). *Looking for Spinoza: Joy, sorrow, and the feeling brain*: Houghton Mifflin Harcourt.

Damasio, A. R., Grabowski, T. J., Bechara, A., Damasio, H., Ponto, L. L., Parvizi, J., & Hichwa, R. D. (2000). Subcortical and cortical brain activity during the feeling of self-generated emotions. *Nature neuroscience, 3*(10), 1049-1056.

Ding, M., Bressler, S. L., Yang, W., & Liang, H. (2000). Short-window spectral analysis of cortical event-related potentials by adaptive multivariate autoregressive modeling: data preprocessing, model validation, and variability assessment. *Biological cybernetics, 83*(1), 35-45.

Eckberg, D. L., & Sleight, P. (1992). *Human baroreflexes in health and disease*: Oxford University Press.

Esslen, M., Pascual-Marqui, R., Hell, D., Kochi, K., & Lehmann, D. (2004). Brain areas and time course of emotional processing. *Neuroimage, 21*(4), 1189-1203.

Fischl, B. (2012). FreeSurfer. *Neuroimage, 62*(2), 774-781.

Fukushima, H., Terasawa, Y., & Umeda, S. (2011). Association between interoception and empathy: evidence from heartbeat-evoked brain potential. *International Journal of Psychophysiology, 79*(2), 259-265.

Granger, C. W. (1988). Some recent development in a concept of causality. *Journal of econometrics, 39*(1-2), 199-211.

Gray, M. A., Taggart, P., Sutton, P. M., Groves, D., Holdright, D. R., Bradbury, D., . . . Critchley, H. D. (2007). A cortical potential reflecting cardiac function. *Proceedings of the National Academy of Sciences, 104*(16), 6818-6823.

Gu, X., Hof, P. R., Friston, K. J., & Fan, J. (2013). Anterior insular cortex and emotional awareness. *Journal of Comparative Neurology, 521*(15), 3371-3388.

Hämäläinen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J., & Lounasmaa, O. V. (1993). Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of modern Physics, 65*(2), 413.

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Harrison, B. J., Pujol, J., Ortiz, H., Fornito, A., Pantelis, C., & Yücel, M. (2008). Modulation of brain resting-state networks by sad mood induction. *PLoS One, 3*(3), e1794.

Harrison, N. A., Gray, M. A., Gianaros, P. J., & Critchley, H. D. (2010). The embodiment of emotional feelings in the brain. *Journal of Neuroscience, 30*(38), 12878-12884.

Hyvärinen, A., Karhunen, J., & Oja, E. (2004). *Independent component analysis* (Vol. 46): John Wiley & Sons.

James, W. (1884). What is an emotion? *Mind, 9*(34), 188-205.

Jeong, B., Lee, S. W., Lee, J.-S., Yoo, J. H., Kim, K. W., Cho, S., . . . Choi, J. (2015). The psychometric properties of the Korean version of the Verbal Abuse Questionnaire in university students. *Psychiatry investigation, 12*(2), 190-196.

Joo, E.-J., Joo, Y.-H., Hong, J.-P., Hwang, S., Maeng, S.-J., Han, J.-H., . . . Kim, Y. S. (2004). Korean version of the diagnostic interview for genetic studies: Validity and reliability. *Comprehensive Psychiatry, 45*(3), 225-229.

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The Phq-9. *Journal of general internal medicine, 16*(9), 606-613.

Luft, C. D. B., & Bhattacharya, J. (2015). Aroused with heart: modulation of heartbeat evoked potential by arousal induction and its oscillatory correlates. *Scientific reports, 5*, 15717.

Müller, L. E., Schulz, A., Andermann, M., Gäbel, A., Gescher, D. M., Spohn, A., . . . Bertsch, K. (2015). Cortical Representation of Afferent Bodily Signals in Borderline Personality Disorder: Neural Correlates and Relationship to Emotional Dysregulation. *JAMA psychiatry, 72*(11), 1077-1086.

Maister, L., Tang, T., & Tsakiris, M. (2017). Neurobehavioral evidence of interoceptive sensitivity in early infancy. *eLife, 6*.

Medford, N., & Critchley, H. D. (2010). Conjoint activity of anterior insular and anterior cingulate cortex: awareness and response. *Brain Structure and Function, 214*(5-6), 535-549.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational intelligence and neuroscience, 2011*, 1.

Pan, J., & Tompkins, W. J. (1985). A real-time QRS detection algorithm. *IEEE transactions on biomedical engineering*(3), 230-236.

Park, H.-D., Bernasconi, F., Bello-Ruiz, J., Pfeiffer, C., Salomon, R., & Blanke, O. (2016). Transient Modulations of Neural Responses to Heartbeats Covary with Bodily Self-Consciousness. *Journal of Neuroscience, 36*(32), 8453-8460.

Park, H.-D., Bernasconi, F., Salomon, R., Tallon-Baudry, C., Spinelli, L., Seeck, M., . . . Blanke, O. (2017). Neural Sources and Underlying Mechanisms of Neural Responses to Heartbeats, and their Role in Bodily Self-consciousness: An Intracranial EEG Study. *Cerebral Cortex*, 1-14.

Park, H.-D., Correia, S., Ducorps, A., & Tallon-Baudry, C. (2014). Spontaneous fluctuations in neural responses to heartbeats predict visual detection. *Nature neuroscience, 17*(4), 612-618.

Park, J. Y., Oh, J. M., Kim, S. Y., Lee, M., Lee, C., Kim, B. R., & An, S. (2011). Korean facial expressions of emotion (KOFEE). *Seoul, Korea: Section of Affect & Neuroscience, Institute of Behavioral*

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Science in Medicine, Yonsei University College of Medicine.

Paulus, M. P., & Stein, M. B. (2010). Interoception in anxiety and depression. *Brain Structure and Function, 214*(5-6), 451-463.

Penny, W. D., Friston, K. J., Ashburner, J. T., Kiebel, S. J., & Nichols, T. E. (2011). *Statistical parametric mapping: the analysis of functional brain images*: Academic press.

Pohl, A., Anders, S., Schulte-Rüther, M., Mathiak, K., & Kircher, T. (2013). Positive facial affect–an fMRI study on the involvement of insula and amygdala. *PLoS One, 8*(8), e69886.

Pollatos, O., Kirsch, W., & Schandry, R. (2005). Brain structures involved in interoceptive awareness and cardioafferent signal processing: a dipole source localization study. *Human brain mapping, 26*(1), 54-64.

Pollatos, O., & Schandry, R. (2004). Accuracy of heartbeat perception is reflected in the amplitude of the heartbeat-evoked brain potential. *Psychophysiology, 41*(3), 476-482.

Seth, A. K., Barrett, A. B., & Barnett, L. (2015). Granger causality analysis in neuroscience and neuroimaging. *Journal of Neuroscience, 35*(8), 3293-3297.

Shao, S., Shen, K., Wilder-Smith, E. P., & Li, X. (2011). Effect of pain perception on the heartbeat evoked potential. *Clinical neurophysiology, 122*(9), 1838-1845.

Singh-Bains, M. K., Waldvogel, H. J., & Faull, R. L. (2016). The role of the human globus pallidus in Huntington's disease. *Brain Pathology, 26*(6), 741-751.

Smith, R., & Lane, R. D. (2015). The neural basis of one's own conscious and unconscious emotional states. *Neuroscience & Biobehavioral Reviews, 57*, 1-29.

Sridharan, D., Levitin, D. J., & Menon, V. (2008). A critical role for the right fronto-insular cortex in switching between central-executive and default-mode networks. *Proceedings of the National Academy of Sciences, 105*(34), 12569-12574.

Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., & Leahy, R. M. (2011). Brainstorm: a user-friendly application for MEG/EEG analysis. *Computational intelligence and neuroscience, 2011*, 8.

Terhaar, J., Viola, F. C., Bär, K.-J., & Debener, S. (2012). Heartbeat evoked potentials mirror altered body perception in depressed patients. *Clinical neurophysiology, 123*(10), 1950-1957.

Vijayaraghavan, L., Vaidya, J. G., Humphreys, C. T., Beglinger, L. J., & Paradiso, S. (2008). Emotional and motivational changes after bilateral lesions of the globus pallidus. *Neuropsychology, 22*(3), 412.

Wang, X., Chen, Y., & Ding, M. (2008). Estimating Granger causality after stimulus onset: a cautionary note. *Neuroimage, 41*(3), 767-776.

Williams, L. M., Das, P., Liddell, B., Olivieri, G., Peduto, A., Brammer, M. J., & Gordon, E. (2005). BOLD, sweat and fears: fMRI and skin conductance distinguish facial fear signals. *Neuroreport, 16*(1), 49-52.

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplementary materials

Table S1. Inclusion criteria of Granger causality ROI

ROI name	MNI coordinates (x, y, z)
RAI	35, 18, 4
RACC	9, 40, 10

ROI was selected from peak voxel of cluster, cluster was defined by uncorrected p value <0.01 and adjacent 10 voxels in sad face – neutral face difference. RAI included 150 and RACC included 25 voxels (Because RAI included more than 2 peak voxels).

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table S2. Clusters showing significant increase of HEP in sad face compared to neutral face condition in absolute time courses (paired-t test)

Cluster	Region	MNI coordinate (x, y, z)	Cluster size (voxel number)	T value	Cluster P value (FWE corrected)
Cluster 1	Rt. vmPFC/Lt OFC	5, 64, 0	729	3.55	0.026
		-26, 54, -11		3.48	
		-26, 60, -17		3.44	

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table S3. Clusters showing significant increase of VEP in sad face compared to neutral face condition in early time range (73ms-198ms, paired-t test)

Cluster	Region	MNI coordinate (x, y, z)	Cluster size (voxel number)	T value	Cluster P value (FWE corrected)
Cluster 1	Rt. middle temporal gyrus/fusiform face area	43, -18, -13	1475	4.77	<0.001
		31, -34, -15		4.67	
		51, -14, -18		4.36	
Cluster 2	Rt. angular gyrus/superior temporal gyrus	57, -56, 26	488	4.25	0.043
		51, -58, 18		3.85	
		45, -52, 12		3.54	

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table S4. Clusters showing significant increase of VEP in sad face compared to neutral face condition in late time range (815ms-948ms, paired-t test)

Cluster	Region	MNI coordinate (x, y, z)	Cluster size (voxel number)	T value	Cluster P value (FWE corrected)
Cluster 1	Rt. supplementary motor area	-5, -6, 58	787	4.62	0.003
		-7, 10, 70		4.36	
		-17, -2, 72		4.01	

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Influence of mood on the HEP modulation effect in happy/sad faces

Because previous studies reported that a depressive mood can affect a person's interoceptive processing, we suspected that there might be a relationship between HEP modulation of emotional expression and mood state. To answer this question, we performed a correlation analysis between the HEP modulation effect of sad/happy faces, which showed a significant HEP modulation at least in the ROI-TOI mask analysis and PHQ-9 score, which represents a depressive state of a person. Because psychological scores typically are not distributed normally, we conducted a Spearman's rank correlation coefficient test between the averaged HEP within the ROI-TOI mask and PHQ-9 score. The HEP modulation was significantly correlated with participants' PHQ-9 score (Spearman's rho = 0.377, $p = 0.033$) for happy faces, while HEP modulation of sad faces was only marginally correlated with PHQ-9 (Spearman's rho = 0.297, $p = 0.099$). Subsequently, we divided participants into two groups – one group who were non or minimally depressive (PHQ-9 score < 5, $n = 19$) and the other group who were mildly depressive (PHQ-9 score between 5 and 9, $n = 13$). This classification of groups was according to the severity classification of PHQ-9. Then we performed a paired t-test between the HEP of a happy face and a neutral face within each group (HEP values were averaged within the ROI-TOI mask). The HEP modulation effect was more significant in participants with a low PHQ-9 score ($t(18) = -3.314$, $p = 0.004$) than in the original statistics from all the participants ($t(31) = -2.12$, $p = 0.042$), while participants with a high PHQ-9 score did not show the HEP modulation effect ($t(12) = 0.30$, $p = 0.766$).

Additional analysis of the HEP modulation effect of a sad emoticon in the sensor level

We tried to determine the reason why a sad emoticon, which is also a sad emotional expression

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

like a sad face, did not modulate HEP although the emotionality scores of a sad face and a sad emoticon were not different (in paired-t test, p = 0.492, mean emotionality score of sad emoticon = 2.81, mean emotionality score of sad face = 2.71). Unlike a face, a text-based emoticon is not an innately encoded stimulus, and responses to sad emoticons are likely to be acquired through experiences, and responses to emoticons may be affected by various factors including, personal experience, which result in variable responses toward emoticons across the population. In our emotionality score, participants showed a more heterogeneous response toward sad emoticons than toward sad faces. The mean variance emotionality score in the sad emoticon across subjects was 1.18, which is much larger than the variance of the sad face (in t (31) = 2.153, in paired t-test). Similarly, we hypothesized that a HEP modulation of a sad emoticon could be more variable relative to HEP modulation of a sad face and more likely to be affected by other factors, like an individual's emotional experience or their current mood state. To prove this hypothesis, first, we tried to determine the psychological trait or state that can modulate the HEP effect of a sad emoticon. To do this, we performed linear regression with HEP modulation by a sad emoticon as a dependent variable and psychological variables with independent variables. Cluster based permutation regression was done using FieldTrip functions 'ft_statfun_indepsamplesregrT'. We hypothesized that the HEP effect would occur in a time window similar to a sad face, while the spatial pattern could be different based on our previous study of emoticon. In that study, regions activated by emotional emoticons were not completely same as typical regions that are activated by emotional face.

Therefore, the HEP values of time windows in which the HEP modulation effect was found for a sad face (488ms-515ms post R peak) were averaged while all sensors were tested and not averaged. Psychological variables used as independent variables included the PHQ-9 score, which reflects participants' mood states and the VAQ score, which corresponds to a (negative) emotional experience. The VAQ included two subscales composed of peer verbal abuse history

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

and parental verbal abuse history. Note that because two participants failed to complete the VAQ, only thirty people were analyzed.

In the results, only the peer verbal abuse score revealed a significant relationship with the HEP difference in the left central sensors (Monte-Carlo p = 0.015). Other psychological variables did not form significant clusters.

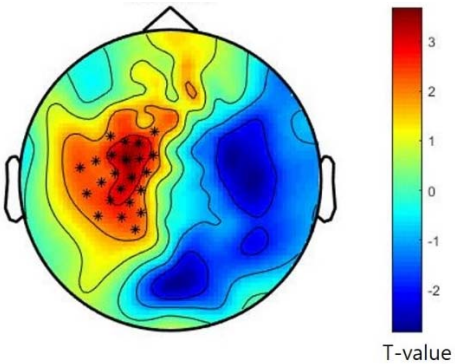

Figure S1. Topography of correlation between PeVA and HEP modulation by sad emoticon.

Based on our finding that HEP modulation is affected by peer verbal abuse history, we classified participants by two groups, the participants in the first group (n = 10) were those who had no peer verbal abuse at all, which correspond to a PeVA score of 15, while the second group consisted of those who had a peer verbal abuse history at least one time (n = 20). Then we performed the paired t–test, which compared the averaged HEP of sad emoticons and neutral emoticons within significant clusters found in the regression analysis for each group. A paired t-test was also done using the Monte-Carlo method because the number of participants in the first group was quite small. The result of the paired t-test showed that there was a significant difference between the HEP of a sad emoticon and a neutral emoticon in the first group, which had no history of verbal abuse (Monte-Carlo p = 0.004), while there was no difference in the second group, which had a verbal abuse history (Monte-Carlo p = 0.212).

bioRxiv preprint doi: https://doi.org/10.1101/200980; this version posted January 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Additionally, we performed the same regression analysis with the happy emoticon as a dependent variable and PeVA, PaVA, PHQ-9 as independent variables. No psychological scores formed clusters that show a significant relationship with HEP modulation of the happy emoticon.

Additional pairwise GC analysis including RPI, RAI and RACC

Although RPI did not show a robust HEP modulation effect, we speculated that there could be a modulation of GC between RPI and RAI by sad face which reflects the modulation of primary interoceptive processing. Therefore, we selected the RPI as an additional ROI, extracted time course from RPI and performed pairwise GC analysis between three ROIs consisting of RPI, RAI, and RACC. MNI coordinate of RPI (x = 39, y = -20, z = 14) was selected from the peak voxel in the cluster showed when we lowered cluster forming threshold in source analysis. Six pairs of GCs including RPI to RAI, RPI to RACC, RAI to RACC, RACC to RAI, RACC to RPI and RAI to RPI were estimated and compared between sad and neutral face condition in the same way as we did in original GC analysis. Similar to our original results, RAI to RACC was significantly increased in sad face (Monte-Carlo p = 0.032) while other GCs including RPI to RAI were not differ between sad and neutral face conditions (RPI to RAI, RAI to RPI and RACC to RPI: did not form clusters, RPI to RACC: Monte-Carlo p = 0.534, RACC to RAI: Monte-Carlo p = 0.706).