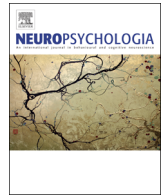




ELSEVIER

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia

The role of prosody and context in sarcasm comprehension: Behavioral and fMRI evidence



Tomoko Matsui^{a,1}, Tagiru Nakamura^{b,1}, Akira Utsumi^c, Akihiro T. Sasaki^{d,g,h},
Takahiko Koike^d, Yumiko Yoshida^{d,e}, Tokiko Harada^d, Hiroki C. Tanabe^{d,f},
Norihiro Sadato^{d,e,*}

^a Center for Research in International Education, Tokyo Gakugei University, 4-1-1 Nukuikitamachi, Koganei, Tokyo 184-8501, Japan

^b Faculty of Environment and Information Studies, Keio University, 5322 Endo, Fujisawa, Kanagawa 252-8520, Japan

^c Department of Informatics, Graduate School of Informatics and Engineering, The University of Electro-Communications, 1-5-1 Chofugaoka, Chofu, Tokyo 182-8585, Japan

^d Department of System Neuroscience, Division of Cerebral Integration, National Institute for Physiological Sciences (NIPS), 38 Nishigonaka, Myodaiji, Okazaki, Aichi 444-8585, Japan

^e Department of Physiological Sciences, School of Life Science, SOKENDAI (The Graduate University for Advanced Studies), Shonan Village, Hayama, Kanagawa 240-0193, Japan

^f Department of Social Environment, Graduate School of Environmental Studies, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8601, Japan

^g Pathophysiological and Health Science Team, RIKEN Center for Life Science Technologies, 6-7-3 Minatojima-minamimachi, Chuo-ku, Kobe, Hyogo 650-0047, Japan

^h Department of Physiology, Osaka City University Graduate School of Medicine, 1-4-3 Asahi-machi, Abeno-ku, Osaka 545-8585, Japan

ARTICLE INFO

Article history:

Received 20 August 2015

Received in revised form

29 April 2016

Accepted 29 April 2016

Available online 6 May 2016

Keywords:

Sarcasm

Prosody

Inferior frontal gyrus

Modulation

Incongruity

Functional magnetic resonance imaging

ABSTRACT

A hearer's perception of an utterance as sarcastic depends on integration of the heard statement, the discourse context, and the prosody of the utterance, as well as evaluation of the incongruity among these aspects. The effect of prosody in sarcasm comprehension is evident in everyday conversation, but little is known about its underlying mechanism or neural substrates. To elucidate the neural underpinnings of sarcasm comprehension in the auditory modality, we conducted a functional MRI experiment with 21 adult participants. The participants were provided with a short vignette in which a child had done either a good or bad deed, about which a parent made a positive comment. The participants were required to judge the degree of the sarcasm in the parent's positive comment (praise), which was accompanied by either positive or negative affective prosody. The behavioral data revealed that an incongruent combination of utterance and the context (i.e., the parent's positive comment on a bad deed by the child) induced perception of sarcasm. There was a significant interaction between context and prosody: sarcasm perception was enhanced when positive prosody was used in the context of a bad deed or, vice versa, when negative prosody was used in the context of a good deed. The corresponding interaction effect was observed in the rostro-ventral portion of the left inferior frontal gyrus corresponding to Brodmann's Area (BA) 47. Negative prosody incongruent with a positive utterance (praise) activated the bilateral insula extending to the right inferior frontal gyrus, anterior cingulate cortex, and brainstem. Our findings provide evidence that the left inferior frontal gyrus, particularly BA 47, is involved in integration of discourse context and utterance with affective prosody in the comprehension of sarcasm.

© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Successful comprehension of sarcasm, which we define here as a subcategory of verbal irony that communicates the speaker's

negative or critical attitude, is often characterized by the hearer's recognition of the gap between the semantic content of the utterance and the speaker's communicative intent. Often, sarcastic utterances are statements that are absurdly inadequate or blatantly false in the light of reality or normative expectations. Consider the following example from Wilson and Sperber (2012):

* Corresponding author at: National Institute for Physiological Sciences (NIPS), Myodaiji, Okazaki 444-8585, Japan.

E-mail address: sadato@nips.ac.jp (N. Sadato).

¹ These authors contributed equally to this work.

Sue (to someone who has done her a disservice): I can't thank you enough.

Having heard the utterance, one cannot help wondering why Sue would wish to thank more extensively a person who has done her a disservice. In fact, if the hearer fails to recognize the oddity of what is described in the utterance, he is also likely to fail to appreciate the speaker's critical attitude. The speaker of a sarcastic utterance, who is aware of this possibility, often tries to provide the hearer with rich but implicit clues regarding how the utterance should be interpreted. For example, tone of voice, facial expressions, and gestures such as head shaking are often used by the speaker as clues to indicate that the utterance should be interpreted as sarcastic (Bryant and Fox Tree, 2005). In this way, the speaker of a sarcastic utterance implicitly highlights the contrast between the semantic content of the utterance and what they intend to communicate. According to Wilson and Sperber (2012), by highlighting this contrast, the speaker intends to communicate his dissociative attitude towards the thought expressed in the sarcastic utterance itself. In other words, recognition of the gap between different levels of meaning on the part of the hearer is a crucial first step in understanding the speaker's intentions.

1.1. Recognition of incongruity in sarcasm comprehension

Once the hearer has recognized the incongruity between what he had expected to hear from the speaker, given a certain conversational context, and what he actually heard (i.e. what is described in the utterance), the next phase of sarcasm comprehension, which is geared toward filling the gaps, begins. Pragmatic models of irony comprehension, including sarcasm, are constructed on the basis of how the hearer's recognition of incongruity during on-line comprehension will yield an ultimate understanding of the speaker's attitude and intentions (Gibbs, 1986; Giora, 1997; Kreuz and Glucksberg, 1989; Kumon-Nakamura et al., 1995; Utsumi, 2000).

Psychological studies on irony processing suggest that the incongruity involved in comprehension of sarcasm is multi-layered. Classic or standard models of irony comprehension suggest that detection of the incongruity between the discourse context (i.e., what happened and how the speaker felt about it in the real world) and the statement enables the hearer to understand the speaker's ironical intent (Ackerman, 1983; Colston, 2002; Katz and Lee, 1993; Katz and Pexman, 1997; Kreuz and Glucksberg, 1989; Ivanko and Pexman, 2003). Furthermore, the larger the disparity between the statement and the discourse context, the more condemning is the irony perceived by the hearer (Colston and O'Brien, 2000; Gerrig and Goldvarg, 2000). Others have argued that the salient difference in prosody between ironic and sincere statements is often enough for a hearer to identify the utterance as an instance of irony (Bryant and Fox Tree, 2005; Capelli et al., 1990). In addition, a developmental study of irony comprehension demonstrated that children first use ironic prosody as an effective clue for irony comprehension around the age of 5 years, before they can also make use of discourse context as a clue (Laval and Bert-Erboul, 2005). Finally, a recent study revealed that there is an interaction between discourse context and the speaker's tone of voice in sarcasm comprehension. According to Woodland and Voyer (2011), when the content of a target utterance was held constant, the combination of the negative discourse context and sarcastic tone of voice was judged most sarcastic, whereas the combination of positive discourse context and sincere tone of voice was least sarcastic, i.e., most sincere. On the other hand, when the discourse context and prosody were incongruent (e.g., when negative discourse context and sincere tone of voice are combined), the utterances were judged neither sarcastic nor sincere, i.e., somewhere in the middle.

In this study, following classic accounts of sarcasm, we assume that in everyday conversation, the hearer typically perceives an

utterance as sarcastic when he recognizes the incongruity between what he expected to hear in the light of a particular context and what he actually heard. This is probably because the hearer's recognition of the incongruity makes him pay more attention to the attitude of the speaker, which in turn enables him to attribute a critical or sarcastic attitude to the speaker (Kumon-Nakamura et al., 1995; Pexman, 2008). In addition, based on previous findings (Colston, 2002; Colston and O'Brien, 2000; Gerrig and Goldvarg, 2000; Ivanko and Pexman, 2003), we assume that the degree of incongruity between the discourse context and the statement will influence the hearer's understanding of the speaker's attitude and intentions.

1.2. Neural substrates for comprehension of sarcasm

Many theoretical and psychological studies of sarcasm, including those reviewed above, suggest that the hearer's perception of an utterance as being sarcastic often depends on his success in recognizing the incongruity between what he expected to hear and what he actually heard. The hearer's expectation of what he should hear is heavily influenced by the available discourse context and what he actually heard, which consists of what was said (i.e., the semantic content of the utterance) as well as how it was said (i.e., prosodic features accompanying the utterance).

Following Ross (2000), we use the notion of affective prosody to cover both emotional (e.g. happiness, sadness or anger) and attitudinal (e.g. endorsement, criticism, or skepticism) prosody, as distinct from the linguistic prosody of the utterance (e.g., sentence focus, word stress, or as declarative or interrogative speech). Affective prosody used in a sarcastic utterance has often been perceived as a natural cue regarding the negative affect or critical or contemptuous attitude that the speaker intends to communicate (e.g. Shamay-Tsoory et al., 2005; Wilson and Sperber, 2012).

Affective prosody should be distinguished from "sarcastic prosody" or a "sarcastic tone of voice," although their acoustic characteristics may partially overlap (Pell, 2006). Sarcastic prosody is a natural acoustic cue to signal sarcastic intent in speech, even in the absence of any semantic or contextual clues (e.g. Bryant and Fox Tree, 2005; Cheang and Pell, 2008; Rockwell, 2007). On the other hand, affective prosody, unlike sarcastic prosody, does not on its own signal the sarcastic intent of the speaker. In this sense, affective prosody has a more restricted role than sarcastic prosody in comprehension of sarcasm. The main role of affective prosody is to signal a specific attitude or emotion of the speaker that is communicated by the utterance.

In a sarcastic utterance, however, the attitude communicated by affective prosody is likely to be incongruent with the semantic or/and contextual information of the utterance. In other words, each of the three factors involved in sarcastic utterance, namely, discourse context, semantic content, and accompanying affective prosody, is likely to play an important but separate role in the hearer's recognition of incongruity as it relates to sarcasm comprehension. We were interested in investigating how incongruity among these three factors influences sarcasm perception, and in elucidating the underlying neural mechanisms.

Previous studies of the neural substrates of sarcasm comprehension suggested that the medial prefrontal cortex (MPFC) and the left inferior frontal gyrus (IFG) are centrally involved (Uchiyama et al., 2006; Wang et al., 2006a; Spotorno et al., 2012). The MPFC is considered to be the main neural basis for mentalizing (Spotorno et al., 2012), whereas the left IFG contributes to integration of linguistic information (Uchiyama et al., 2006). More generally, the left IFG is widely believed to play a crucial role in processing semantic integration and evaluation (Dapretto and Bookheimer, 1999; Gabrieli et al., 1996; Kapur et al., 1994; Rapp

et al., 2004; Uchiyama et al., 2006; Wagner et al., 1997). The left IFG, spanning from Brodmann's Area (BA) 47 to BA 44, is regarded as a "unification space" with a functional gradient, oriented in a rostro-ventral to caudo-dorsal direction, that enables integration of semantics, syntax, and phonology during sentence comprehension (Hagoort, 2005). Furthermore, probably through the connection with the superior temporal gyrus (STG), the bilateral IFG is involved in evaluation of affective tone of voice, which is relevant for the executive processes such as controlling, overriding, or inhibiting behavioral and emotional responses (Frühholz and Grandjean, 2013). Therefore, although the neural substrates of the hearer's recognition of potential incongruity between what he expected to hear and what he actually heard during comprehension of sarcasm have not been previously investigated, it seems reasonable to hypothesize that the left IFG is also involved in integration and evaluation of semantic content, discourse context, and affective prosody of sarcastic utterances, ultimately yielding recognition of the incongruity among them.

Importantly, previous investigations of neural substrates of sarcasm comprehension typically used only written sarcasm stimuli. As a result, the neural mechanisms that underlie processing of the affective prosody accompanying a sarcastic utterance remain virtually unknown. Of the previous neuroimaging studies on comprehension of sarcasm, only two included prosodic information among the experimental stimuli (Wang et al., 2006a, 2006b). Moreover, in those studies, a typical sarcastic or negative tone of voice was always paired with negative discourse context (sarcasm), and the typical sincere or positive tone of voice was always paired with positive discourse context (sincere praise). In other words, no attempt was made to separate the neural basis for processing affective prosody from the neural activities involved in processing of discourse context during sarcasm comprehension. Thus, the neural basis for the detection of incongruity between discourse context and the statement the hearer actually heard, which consists not only of its semantic content but also of its affective tone, has not been previously investigated.

To date, the neuro-psychological mechanism underlying processing of a speaker's attitude conveyed via prosody (as in the case of sarcasm in everyday conversation) has been investigated far less extensively than the processing of a speaker's vocal emotion (Mitchell and Ross, 2013). Existing neuro-cognitive models of affective prosody comprehension focus solely on emotional prosody (e.g. Frühholz and Grandjean, 2013; Schirmer and Kotz, 2006; Wildgruber et al., 2006), and potential neuroanatomical differences between processing of emotional and attitudinal prosody are yet to be explored. Given the paucity of studies on processing of attitude conveyed by prosody in general, in the following we will discuss neuro-psychological models of emotional prosody processing and suggest how these models can be used as a starting point for the first study aimed at identifying the role of prosody in sarcasm comprehension.

Previous neuroimaging studies demonstrated the involvement of several neural areas other than auditory cortex in processing of vocal emotion (Ethofer et al., 2012). For example, according to a three-stage model suggested by Schirmer and Kotz (2006), the first sensory processing stage involves bilateral auditory processing areas, whereas the second integration stage engages the superior temporal gyrus and the anterior superior temporal sulcus. During the last stage of cognition stage in the model, explicit evaluative judgments of affective prosody are mediated by the right IFG and the orbitofrontal cortex, whereas the integration of affective prosody into language processing recruits the IFG in the left hemisphere. Based on other neuroimaging results, Wildgruber et al. (2009) proposed a similar model in which the first step, bottom-up modulation or extraction of supra-segmental acoustic information, is associated predominantly with activation of the

right hemispheric primary and secondary acoustic regions including the right mid superior temporal cortex, which is specifically responsive to human voices (Belin et al., 2000). The second step, representation of meaningful supra-segmental acoustic sequences, is linked to posterior aspects of the right superior temporal sulcus (STS). The third step, emotional judgment, is linked to the IFG. Connectivity analysis of cerebral activation revealed that the right post-STG is the most likely input region into the network of areas characterized by task-dependent activation (Ethofer et al., 2006). Therefore, the right post-STG subserves the representation of meaningful prosodic sequences and receives direct input from primary and secondary acoustic regions. Because the right STG and the bilateral IFG are activated by the explicit judgment of the emotional valence conveyed by prosody (Ethofer et al., 2006), the second and third steps are likely to represent the process dependent on focusing of attention towards explicit emotional evaluation (top-down effects). Both models indicate the task-dependent involvement of the IFG during evaluation of affective prosody. In addition, Frühholz and Grandjean (2013) suggested that the bilateral IFG is involved in executive processes such as evaluation and categorization of emotional prosody provided by higher-level auditory regions in the STG. Although confirming that the right IFG is predominant in cognitively controlled evaluation of emotional prosody, they argued against the existing view that involvement of the left IFG is restricted to language-related features of emotional prosody. They suggested instead that the left IFG plays more generic roles in processing of emotional prosody, e.g. evaluation and categorization. On the basis of these models, we suggest that the left IFG is a likely candidate for the neural basis of integration of affective prosody into language processing during sarcasm comprehension.

Three previous studies investigated the neural basis of integration of affective prosody and linguistic information during speech perception, although they did not focus on sarcasm comprehension per se. Using functional magnetic resonance imaging (fMRI), Schirmer et al. (2004) compared brain regions that mediate processing of two types of emotional speech: compatible (e.g., positive word with happy voice) or incompatible (e.g., positive word with angry voice) combinations of emotional prosody and word meaning. The results revealed that the left IFG was more strongly activated during the processing of incompatible stimuli (e.g., the combination of positive word meaning and angry prosody) than compatible stimuli. Similarly, Mitchell (2006) compared functional brain responses to three different types of utterances: (a) those with compatible semantic content and emotional prosody; (b) those with incompatible semantic content and emotional prosody; and (c) those with emotional prosody and low-pass-filtered semantic content (prosody-only condition). The results suggested that the left IFG, bilateral superior and middle temporal gyri, and basal ganglia are associated with processing of utterances with incompatible combinations of semantic content and emotional prosody. More recently, Wittfoth et al. (2010) reported that the left IFG and the middle temporal gyrus are engaged in processing of utterances with happy prosody and negative semantic content. Collectively, these studies demonstrated that the left IFG is centrally involved in processing incompatible combinations of emotional prosody and semantic content.

These findings indicate that the left IFG contributes not only to integration of emotional prosody and semantic information of the same utterance, but also to detection of valence compatibility between them. Therefore, in this study, we addressed the novel question of whether this function of the left IFG also extends to integration of affective prosody and semantic content of the utterance in sarcasm comprehension.

1.3. Current study: the role of affective prosody in the comprehension of sarcasm

The hearer's perception of an utterance as being sarcastic often depends on his detection of the incongruity between what is indicated by the discourse context and the statement he actually heard. In everyday conversation, the statement the hearer actually heard consists not only of its semantic content, but also of a particular affective prosody accompanying it. Thus, before the hearer recognizes the incongruity between the discourse context and the statement, what is indicated by the particular affective prosody needs to be integrated with semantic content. Therefore, the goal of this study was to investigate, for the first time, the neural substrates of the integration of affective prosody with the meaning of an utterance during the process leading to incongruity detection in comprehension of sarcasm.

Adopting standard accounts of sarcasm comprehension (Ackerman, 1983; Colston, 2002; Katz and Pexman, 1997; Kreuz and Glucksberg, 1989; Ivanko and Pexman, 2003), here we assume that the semantic content of the utterance and the accompanying affective prosody are considered as part of what the hearer actually heard. In other words, both the semantic content and affective prosody of the utterance together contribute to the hearer's perception of what the speaker intended to communicate. Therefore, when there is potential incongruity between what the hearer expected to hear on the basis of the context available and what he actually heard, the degree of incongruity is evaluated solely between the overall meaning of the utterance and the discourse context. That is, neither the semantic content of the utterance alone, nor the accompanying affective prosody alone, is evaluated during incongruity detection in sarcasm comprehension.

We also assume that evaluation of incongruity between the statement and its discourse context consists of two stages. In the first stage, 'integration by modulation,' affective prosody is integrated into language processing during which it modulates the positive or negative valence of semantic content. For example, when positive prosody accompanies an utterance with positive semantic content (compatible combination), the overall positive valence of the utterance meaning is strengthened. By contrast, negative prosody accompanying an utterance with positive semantic content would reduce the overall positive valence of the utterance meaning. In the second stage, 'evaluation,' overall utterance meaning (combination of semantic content and affective prosody), which is the end product of the modulation stage, is evaluated for its congruity or compatibility with the discourse context.

In order to examine the neural areas involved in incongruity evaluation in sarcasm comprehension, we compared four experimental conditions. Crucially, across conditions, the semantic content of the target utterance was kept positive. The remaining two factors, namely, affective prosody and discourse context, were categorized into simple binary opposites of either positive or negative in each condition. The resulting four conditions were as follows: the combination of bad behavior (as the discourse context) and negative prosody (BN); bad behavior and positive prosody (BP); good behavior and positive prosody (GP); and good behavior and negative prosody (GN).

Furthermore, on the basis of standard theories of sarcasm comprehension, we predicted that the resulting degree of incongruity between the discourse context and overall utterance meaning would influence the likelihood that the target utterance would be perceived as being sarcastic. For example, in the BP condition the target utterance would be perceived as more sarcastic than in the BN condition. By contrast, in the GP condition the target utterance would be perceived as less sarcastic than in the GN condition. We collected behavioral data to test these predictions.

On the basis of existing models of speech prosody (Ethofer et al., 2006; Frühholz and Grandjean, 2013; Wildgruber et al., 2009) and previous findings that the left IFG is centrally involved in processing linguistic input whose semantic content is incompatible with accompanying affective prosody (Mitchell, 2006; Schirmer et al., 2004; Wittfoth et al., 2010), we hypothesized that the effect of prosody on sarcasm perception is expected in the left IFG. Specifically, we predicted that activation of the left IFG would be increased when the valence of affective prosody conflicts with a discourse context that enhances sarcasm perception.

2. Methods

2.1. Participants

Twenty-four participants were recruited as paid volunteers for the fMRI experiment, but three participants were excluded due to high rates of response errors in the judgment phase, leaving 21 participants for the final analysis (13 females and 8 males; mean age, 20.5 years; range, 19–27 years). All participants had normal or corrected-to-normal visual acuity and were right-handed (mean score: 87.7; range, 51.5–100) according to the Edinburgh handedness inventory (Oldfield, 1971); no history of neurological or psychiatric illness was identified. Written informed consent to participate in this study was obtained following procedures approved by the Ethical Committee of the National Institute for Physiological Sciences, Japan.

2.2. Preparation of task materials

In order to examine how affective prosody contributes to recognition of incongruity between discourse context and utterance meaning in sarcasm comprehension, we used a 2×2 factorial design with discourse context (positive or negative) and affective prosody (positive or negative) of the target utterance as independent variables and positive semantic content as a dependent variable. Discourse context depicted either a good or bad deed of the protagonist, and the target utterance was a positive comment on that deed. We assumed that affective prosody plays a role in modulating the semantic valence of the target utterance, i.e., when the utterance was accompanied by positive prosody, its positive valence would be strengthened; by contrast, when the target utterance was accompanied by negative prosody, its positive valence would be reduced.

As experimental stimuli, we used a set of daily conversations between parent and child. Each stimulus consisted of four distinctive phases (Fig. 1), in which were presented: (1) the relevant background situation; (2) the parent's utterance to the child; (3) the child's reaction to the parent's utterance; and (4) the parent's sarcastic or sincere comment about the child's reaction. The first three phases were demonstrated using illustrations (drawn by a professional illustrator to reduce situational ambiguity) as well as written texts. In the fourth phase, the parent's comments were presented either aurally or via written text in an illustration. When the comments were aurally presented, they were accompanied with either negative or positive affective prosody (recorded by a professional actor and actress). An example of mother–child conversation is given below:

- (1) *A boy was playing with lots of toys.*
- (2) *His mother told him that he should tidy the toys before having his snack.*
- (3) *The boy started eating his snack before tidying the toys.*
- (4) *His mother said, "You did a great job tidying the toys!"*

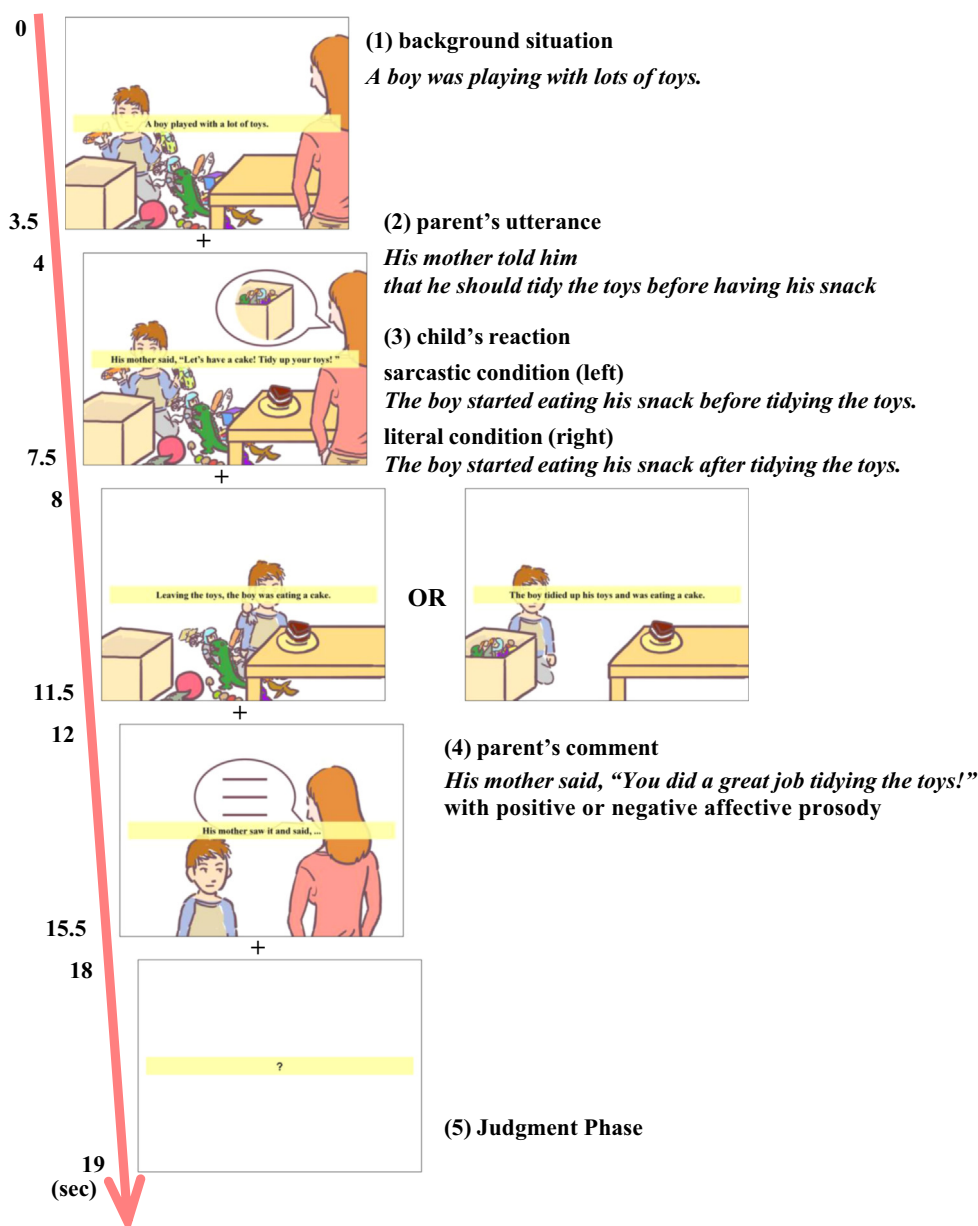


Fig. 1. Time course of an experimental trial: (1) the first phase introduces the relevant background situation; (2) the second phase presents the parent's utterance to the child; (3) the third phase shows the child's reaction to the parent's utterance; and (4) the fourth phase presents the parent's sarcastic or sincere comment about the child's reaction. In the judgment phase (J), participants judged whether the speaker really meant what she/he said.

In this example, in the third phase, the child did not follow the mother's instruction given in the second phase. In this context, the mother's comment in the fourth phase should be interpreted as an example of sarcasm. By contrast, in the following example, in which the child behaves as expected in the third phase, the same comment given by the mother in the fourth phase should be interpreted as a sincere praise.

- (1) A boy was playing with lots of toys.
- (2) His mother told him that he should tidy the toys before having his snack.
- (3) The boy started eating his snack after tidying the toys.
- (4) His mother said, "You did a great job tidying the toys!"

In order to investigate the neural basis for processing affective prosody sarcasm comprehension, we created four experimental conditions, each of which is represented by an abbreviation as

defined here: the BN condition (the child behaved *badly* in the third phase [B] and the parent gave a comment on his behavior with *negative* affective prosody in the fourth phase [N]); the BP condition (the child behaved *badly* in the third phase [B] and the parent gave a comment on his behavior with *positive* affective prosody in the fourth phase [P]); the GN condition (the child's behavior was *good* in the third phase [G] and the parent gave a comment on his behavior with *negative* affective prosody in the fourth phase [N]); and the GP condition (the child's behavior was *good* in the third phase [G] and the parent gave a comment on his behavior with *positive* affective prosody in the fourth phase [P]).

We included two filler conditions, which are also represented by abbreviation: the BW filler condition (the child behaved *badly* in the third phase, and the parent gave a comment expressed by a *written* script in the fourth phase), and the GW filler condition (the child's behavior in the third phase was *good* and the parent gave a comment expressed by a *written* script in the fourth phase). We

Table 1

Acoustical analyses for all conditions of affective prosody: duration mean, intensity mean, and F0 mean.

	Conditions for affective prosody									
	N		P		P		A			
Affective prosody	N		N		P		P		Reversed	
Semantic content	Praise		Praise		Praise		Praise		Nonsense	
Actor	F		M		F		M		F and M	
Duration (sec)	2.4	(0.19)	2.0	(0.29)	2.3	(0.15)	2.5	(0.22)	2.4	(0.19)
Intensity (dB)	73.6	(0.41)	69.8	(0.96)	75.0	(1.16)	68.7	(1.14)	69.5	(1.31)
F0 (Hz)	195.1	(7.6)	112.8	(3.2)	257.2	(8.4)	107.6	(3.0)	110.3	(4.5)

Abbreviations: N/P, parent's negative/positive affective prosody in the fourth phase; A, auditory baseline (nonsense sound created by reversing a parent's utterance) in the fourth phase; F/M, female/male actors.

also added two control conditions (four trials each); an auditory control condition (A), in which reversed parents' utterances and no written letters were presented in the fourth phase, and a visual control condition (V), in which randomized written letters and no auditory input were presented. We prepared a total of 48 trials (eight each for BN, BP, BW, GN, GP, and GW, and four trials each for A and V) administered in four functional runs.

The parent's comments in the fourth phase were recorded by one female and one male actor in a silent room using a microphone (SM58; Shure, Evanston, IL, USA), an audio interface (0202; E-MU Systems, Scotts Valley, CA, USA), and a personal computer (ThinkPad X201; Lenovo, Morrisville, NC, USA). As shown in [Table 1](#), acoustical analyses for all conditions of affective prosody were carried out using the PRAAT software version 5.4 (<http://www.praat.org/>): mean duration (main effect of affective prosody: $F(1, 12)=1.208$, *n.s.*; main effect of female/male: $F(1, 12)=0.134$, *n.s.*; interaction between affective prosody and female/male: $F(1, 12)=2.081$, *n.s.*), mean intensity (main effect of affective prosody: $F(1, 12)=0.024$, *n.s.*; main effect of female/male: $F(1, 12)=27.433$, $p < 0.001$; interaction between affective prosody and female/male: $F(1, 12)=1.640$, *n.s.*), and mean F0 (main effect of affective prosody: $F(1, 12)=22.006$, $p < 0.001$; main effect of female/male: $F(1, 12)=365.761$, $p < 0.001$; interaction between affective prosody and female/male: $F(1, 12)=30.683$, $p < 0.001$). For recordings by the female actor, positive affective prosodies were higher in frequency than negative affective prosodies (simple main effect: $F(1, 24)=52.329$, $p < 0.001$), whereas there was no significant difference between positive and negative affective prosodies in recordings by the male actor (simple main effect: $F(1, 24)=0.360$, *n.s.*).

To determine whether our experimental stimuli with these auditory utterances were indeed perceived as sarcasm in the bad context (BN and BP) conditions, 50 volunteers (26 females and 24 males; mean age, 21.2 years; range, 18–32 years) participated in a norming study. We presented the experimental stimuli in a pseudo-random order and asked the participants whether the parent's comment in the fourth phase was an example of sarcasm, sincere praise, or neither. The mean proportions of sarcasm judgments were: 95.5% for the BN condition, 64.0% for the BP condition, 29.8% for the GN condition, and 3.8% for the GP condition. A two-way ANOVA of discourse context (Bad, Good) and emotional prosody (Positive, Negative) conducted on the angular-transformed proportion of sarcasm judgments revealed that the main effects of discourse context ($F(1, 49)=269.69$, $p < 0.001$) and emotional prosody ($F(1, 49)=81.31$, $p < 0.001$) were significant, whereas the interaction between these two factors was not. These results demonstrate that our experimental stimuli were well controlled, in that praise for the bad deed (BN and BP conditions) was interpreted as sarcasm, whereas praise for the good deed (GN and GP conditions) was interpreted as sincere.

2.3. fMRI procedures

Prior to the fMRI session, participants were given detailed instructions of the task procedure. In order to familiarize participants with the task, they were also provided with examples of stimuli that did not appear during the fMRI session. All stimuli were presented using the Presentation 14.8 software (Neurobehavioral Systems, Albany, CA, USA) running on a personal computer (Dimension 9200; Dell Computer, Round Rock, TX, USA). Using a liquid crystal display (LCD) projector (DLA-M200L; Victor, Yokohama, Japan), the visual stimuli were projected onto a half-transparent viewing screen located behind the head coil of the magnetic resonance imaging (MRI) scanner. Participants viewed the stimuli via a tilted mirror attached to the head coil. The spatial resolution of the projector was 1024×768 pixels, with a 60-Hz refresh rate. The distance between the screen and the eyes of the subjects was approximately 60 cm, and the visual angle was 18.9° (horizontal) \times 14.2° (vertical). Sentence stimuli (maximum visual angle, $16.5^\circ \times 0.9^\circ$) were written in Japanese (the first language of the participants) and presented in black letters (visual angle, $18.9^\circ \times 14.2^\circ$). Auditory stimuli were presented via MR-compatible headphones (Hitachi, Yokohama, Japan).

Each of the four phases of each trial was presented on the screen for 3.5 seconds, followed by a fixation cross on a black screen (visual angle, $0.6^\circ \times 0.6^\circ$) for 0.5 s ([Fig. 1](#)). After the 4 phases, an additional fixation cross was presented for 2 s, and then the participant was required to judge whether what the parent said in the fourth phase was what he/she really wanted to say (sincere praise) or not (sarcasm), and to respond by pressing a button with their right index or middle finger as quickly as possible while the question mark "?" (visual angle, $0.6^\circ \times 0.6^\circ$) appeared on the screen for 1 s. After the participant's response, a fixation cross was shown again on the screen for 6 s.

We used an event-related design to minimize habituation and learning effects. The 48 task trials (8 scenarios \times 2 discourse contexts \times (2 affective prosodies + 1 no-prosody) and eight control trials were presented in a pseudo-random order. The experiment consisted of four runs, each consisting of 12 experimental trials (2 sets of [2 discourse contexts \times (2 affective prosodies + 1 no-prosody)]), one auditory control trial and one visual control trial. The presentation order of the four runs was counterbalanced across participants.

All images were acquired using a 3-Tesla MR scanner (Allegra; Siemens, Erlangen, Germany). An ascending T2*-weighted gradient-echo echo-planar imaging (EPI) procedure was used in functional imaging to produce 34 continuous transaxial slices covering the entire cerebrum and cerebellum (time echo [TE], 30 ms; flip angle, 85° ; field of view [FoV], 192 mm; 64×64 matrix; voxel dimensions, 3.0×3.0 mm in plan, 4.0 mm slice thickness with 15% gap). A "sparse sampling" technique was used to minimize the effects of image acquisition noise on task performance.

Repetition time (TR) between two successive acquisitions of the same slice was 5000 ms. Cluster volume acquisition time was 2000 ms, leaving a 3000-ms silent period in which the sound stimuli in the fourth phase were presented. Oblique scanning was used to exclude the eyeballs from the images. Each run consisted of a continuous series of 75 vol acquisitions, resulting a total duration of 6 min 15 s. For anatomical imaging, T1-weighted magnetization-prepared rapid-acquisition gradient-echo (MP-RAGE) images were also obtained (TR, 2500 ms; TE, 4.38 ms; flip angle, 8°; FoV, 230 mm; 1 slab; number of slices per slab, 192; voxel dimensions, 0.9 × 0.9 × 1.0 mm) for each participant. The total duration of the experiment was around 90 min for each participant.

2.4. Data analysis

2.4.1. Behavioral performance

A two-way ANOVA with two within-subject factors, namely discourse context and emotional prosody, was conducted on the angular-transformed proportions of responses to the question of whether what the parent said in the fourth phase was what he/she really wanted to say (sincere praise) or not (sarcastic praise). The analysis was carried out using SPSS version 22.0 software (IBM, Armonk, NY, USA).

2.4.2. Imaging data

Preprocessing of the imaging data was performed as follows. The first two EPI volumes of each run were discarded due to unstable magnetization, and the remaining 73 EPI volumes per run (a total of 292 EPI volumes per participant) were analyzed using Statistical Parametric Mapping 8 (SPM8; Wellcome Department of Imaging Neuroscience, London, UK; Friston et al., 2007) implemented in MATLAB (Mathworks, Natick, MA, USA). EPI volumes were spatially realigned to correct for head motion. Next, the T1 weighted anatomical image was co-registered to the mean image of the EPI volumes, segmented into gray and white matter, reconstructed (including a procedure for signal inhomogeneity correction), and spatially normalized to the Montréal Neurological Institute T1 template. The normalization parameters of the T1 weighted anatomical image were applied to all the EPI volumes, and then spatially smoothed in three dimensions using an 8 mm full-width half-maximum Gaussian kernel.

After preprocessing, individual analysis of the EPI data obtained for each participant was conducted using a general linear model. The fourth phase of the four experimental conditions (BN, BP, GN, GP), two filler conditions (BW, GW) and two control conditions (A, auditory control condition [reversed utterance in the fourth phase]; V, visual control condition [nonsense written letters and no auditory input in the fourth phase]) were separately modeled by convolution with a hemodynamic response function. The first, second, and third phases were collapsed together and also modeled as a regressor (C, context in the first, second, and third phase, which was modeled out) by convolution with a hemodynamic response function. Additionally, button responses were modeled as an independent regressor (J, judgment phase) using a convolved delta function. High-pass filters (128 s) were applied to the time-series data. An autoregressive model was used to estimate the temporal autocorrelation. The signal of EPI images was scaled to a grand mean of 100 overall voxels and volumes within each run. Six regressors for head movement parameters obtained in the realignment process were entered in the model. To depict the activations evoked from the same control condition, we made the following contrasts: BN versus A (auditory control condition) [BN–A], BP versus A [BP–A], GN versus A [GN–A], and GP versus A [GP–A].

The contrast images, which consisted of the weighted sum of parameter estimates and represented the normalized task-related

increment of the MR signal obtained in the individual analyses, were subjected to group analysis with a random-effects model to make population-level inferences regarding task-related activation. In total, data from 21 participants and four contrasts (BN–A, BP–A, GN–A, and GP–A) were incorporated into the 2 (discourse context) × 2 (affective prosody) within-subject factorial design (Friston et al., 2007). Specifically, using the flexible factorial design model (Friston et al., 2007), a subject factor was set as independent to take different individuals into account. Error variance was set as equal across participants because they were sampled from the same underlying population. On the other hand, two condition factors were set as dependent because the different factor levels were correlated within subject, with equal error variances because they were taken from the same subjects. To show activations related to processing affective prosody in sarcasm comprehension, we created the following contrasts: the interaction contrast [–(BN–A)+(BP–A)+(GN–A)–(GP–A)] (i.e., [–BN+BP+GN–GP]) and the main effect of affective prosody [(+BN–A)–(BP–A)+(GN–A)–(GP–A)] (i.e., [+BN–BP+GN–GP]).

The resulting set of voxel values for each contrast constituted a statistical parametric mapping (SPM) of the *t* statistic, which was transformed into normal distribution units with a threshold set at $Z > 3.09$ ($p < 0.001$) at the voxel level and $p < 0.05$ with a correction for multiple comparisons at the cluster level for the entire brain.

To confirm the involvement of the IFG regarding the main effect of affective (negative) prosody, we evaluated the overlap between the activation clusters and the pre-defined ROIs (bilateral BA 44 and BA 45) provided by SPM Anatomy Toolbox version 2.1 (Eickhoff et al., 2007), using MarsBaR version .44 (<http://marsbar.sourceforge.net>).

3. Results

3.1. Behavioral performance

As expected, when emotional prosody and semantic content were incongruous (negative prosody), in contrast to congruous positive prosody, the percentage of insincerity judgments decreased when the discourse context and the semantic content were incongruous (praise for bad deed), whereas it increased when they were congruous (praise for good deed) (Fig. 2). The proportions of sarcastic responses were 79.8% for the BN condition, 97.6% for the BP condition, 40.5% for the GN condition, and 1.2% for the GP condition (Fig. 2). A two-way ANOVA of discourse context (Bad, Good) and prosody (Positive, Negative) revealed a significant main effect of discourse context, $F(1, 20) = 127.11$, $p < 0.001$, and a significant main effect of prosody, $F(1, 20) = 12.03$, $p < 0.01$. There was a significant interaction between these two factors, $F(1, 20) = 19.50$, $p < 0.001$. The nature of this interaction was such that when a discourse context involved a bad event, an utterance accompanied by positive prosody was judged as significantly more sarcastic than an utterance with negative prosody, $F(1, 40) = 7.17$, $p < 0.05$. By contrast, for a discourse context involving a good event, an utterance with positive prosody was judged as significantly less sarcastic than one with negative prosody, $F(1, 40) = 29.47$, $p < 0.001$. This result indicates that positive prosody facilitates sarcastic interpretation of an utterance with positive semantic valence in a bad context, possibly by enhancing the overall positive valence of the utterance and thereby increasing the degree of incongruity between utterance meaning and discourse context. On the other hand, it inhibits sarcastic interpretation when an utterance with positive semantic valence is used in a positive context, because the enhancement of the positive valence of the utterance eliminates the incongruity between

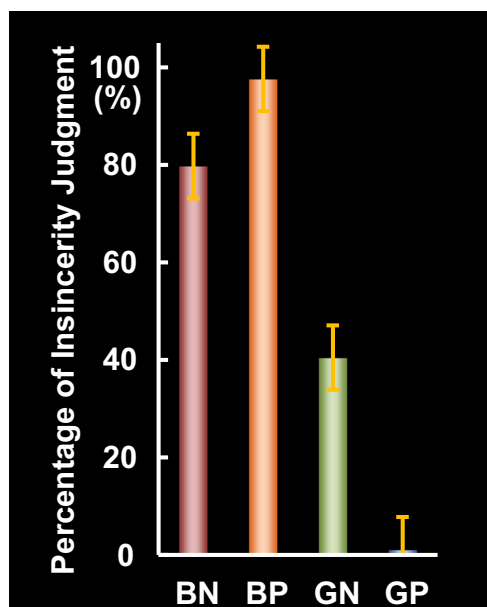


Fig. 2. The percentage of answers indicating that what the parent said in the fourth phase was not what he/she really meant (sarcastic interpretation) were as follows: 79.8% for the BN condition, 97.6% for the BP condition, 40.5% for the GN condition, and 1.2% for the GP condition. When the child exhibited bad behavior in the third phase (B), the degree of incongruity in the BP condition was higher than in the BN condition. When the child exhibited good behavior in the third phase (G), the degree of incongruity in the GN condition was higher than in the GP condition. Abbreviations: B/G, child's bad/good behavior in the third phase; N/P, parent's negative/positive prosody in the fourth phase.

positive utterance meaning and context involving a good event.

3.2. Group analysis of fMRI data

We observed corresponding interaction effects in utterance-related neural activation in the left IFG (BA 47; the Montréal Neurological Institute [MNI] coordinates $[-38\ 34\ -8]$; Fig. 3 (a) and Table 2a). A main effect of negative prosody was found in the bilateral anterior insula (AI; BA 13; $[-30\ 18\ -8]$; $[32\ 22\ -12]$), the posterior rostral zone of anterior cingulate cortex (ACC; BA 32/8; $[2\ 28\ 40]$), and the brainstem ($[0\ -22\ -24]$) (Fig. 3(b) and Table 2b). In the activated cluster (877 voxels) including the left AI, 2.1% (18 voxels) were in left BA 44, and 0.5% (4 voxels) were in left BA 45. In the mass (2482 voxels) including the right AI, 4.8% (118 voxels) were in right BA 44, and 15.2% (377 voxels) were in right BA 45.

4. Discussion

Let us summarize the main findings of the current study. First, analysis of the behavioral data confirmed our predictions concerning perception of sarcasm and interaction of affective prosody and discourse context. When positive prosody was combined with positive semantic content, it enhanced the overall positive valence of utterance meaning. On the other hand, when negative prosody was combined with positive semantic content, the overall positive valence of utterance meaning was reduced. As a result, greater incongruity was perceived in the BP condition than in the BN condition. Consequently, utterances in the BP condition were judged as more sarcastic than those in the BN condition. On the other hand, negative prosody used in a positive discourse context, when combined with positive semantic content as in the GN condition, reduced the positive valence of utterance meaning and thus created incongruity between the utterance meaning and

discourse context. As a result, utterances in the GN condition were perceived as more sarcastic than those in the GP condition.

Regarding the neural correlates, the left rostro-ventral IFG (BA 47) exhibited a significant interaction effect, indicating that incongruent prosody enhances the neural response of the left IFG to the context-dependent perception of the utterance. Thus, the left rostro-ventral IFG may neurally represent the integration of the statement, context, and prosody. On the other hand, the right IFG, including both rostral and caudal portions, was activated by negative prosody. Importantly, in this study, the uttered words were always “praise,” which is semantically positive, whereas the affective prosody was negative or positive. In other words, the main effect of negative prosody observed in this study may represent incongruity detection between the statement (praise) and negative prosody. Therefore, there is a functional asymmetry in terms of incongruity processing by the right (statement-prosody) and the left (statement-prosody-context) IFG.

Previous fMRI studies of comprehension of figurative language and sarcasm reported that the left IFG is activated during comprehension (Rapp et al., 2004, 2010; Spotorno et al., 2012; Uchiyama et al., 2006; Zemleni et al., 2007). Some of these studies suggested that the main function of the left IFG is to integrate semantic information with higher-order mindreading. Other studies claimed that the left IFG is involved in the process of selecting semantic information from a set of competing alternatives (Gabrieli et al., 1996; Petrides, 2005; Sakai, 2005; Thompson-Schill, 2003; Thompson-Schill et al., 1997; Thompson-Schill et al., 1999; Turken and Dronkers, 2011). The contribution of the left IFG, however, may not be tied to semantic processes, but instead may also apply to the selection process in other non-linguistic domains of cognition (Banich et al., 2001; Leung et al., 2000; Mead et al., 2002; Milham et al., 2001; Peterson et al., 2002; Zysset et al., 2001). It has also been suggested that the left IFG plays a key role in integrating world knowledge and sentence contexts (Hagoort et al., 2004; Menenti et al., 2009; Rapp et al., 2011). In light of these findings, it seems likely that the left IFG is involved in both selection and integration of a set of competing information in order to yield an interpretation of what is going on. The results of this study sheds further light on the potential function of the left IFG by demonstrating that this region, particularly BA 47, is also involved in the comprehension of sarcasm. It carries an important function of identifying compatibility among perceived information, integrating different aspects of utterance meaning via modulation on the basis of the identified incompatibility, and evaluating congruity between relevant world knowledge (discourse context) and overall utterance meaning, which assists in perception of sarcasm.

We also observed an effect of statement-prosody incongruity in the ACC and the anterior insula adjacent to the inferior frontal gyrus. It has been suggested that the anterior insula, the right fronto-insular cortex, and the anterior cingulate cortex form a “salience network” that marks salient events and initiates appropriate control signals for additional processing (Menon and Uddin, 2010; Sridharan et al., 2008). In this context, it is conceivable that the salience network, incorporated with the right IFG, is involved in detection of statement-prosodic incongruity that in turn initiates further processing of incorporating context mediated by the rostral IFG (BA 47).

We wish to note that this study had some limitations. For example, in order to keep the total duration of the fMRI experiment within a reasonable interval, we restricted our target utterances to those with positive semantic valence. Future research should investigate whether the neural mechanism underlying recognition of incongruity in sarcasm comprehension identified in this study also applies to statements with negative semantic valence.

Furthermore, several potentially interesting questions

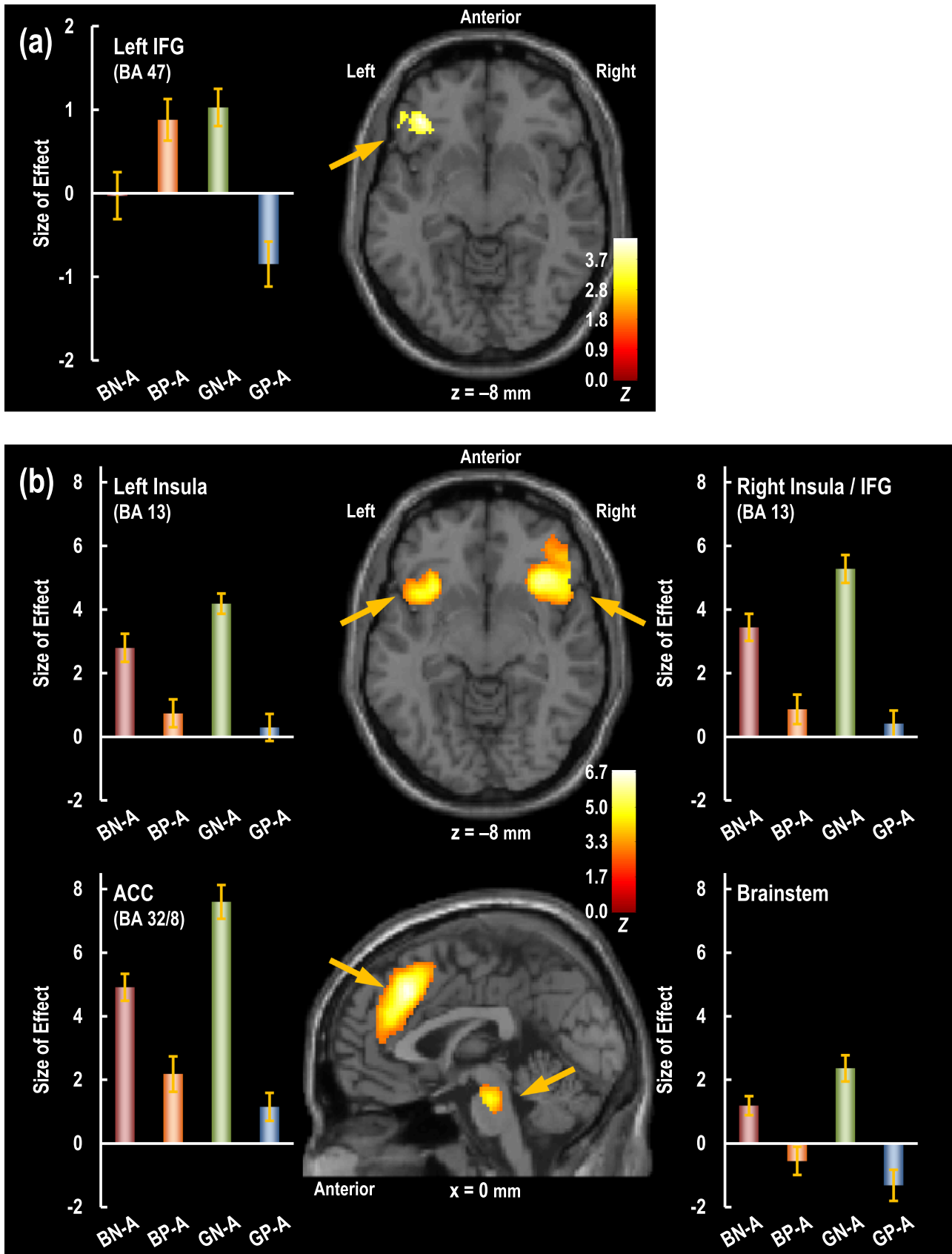


Fig. 3. Activations related to affective prosody in sarcasm comprehension. Each graph was drawn setting the auditory baseline as zero. (a) The left inferior frontal gyrus (IFG; BA 47; MNI coordinates $[-38\ 34\ -8]$) was activated by the interactional contrast of $[-BN + BP + GN - GP]$. (b) The bilateral anterior insula (AI; BA 13; $[-30\ 18\ -8]$; $[32\ 22\ -12]$) extending to the right inferior frontal gyrus (IFG, BA 44, 45, 47) (see Table 2), the anterior cingulate cortex (ACC; BA 32/8; $[2\ 28\ 40]$), and the brainstem ($[0\ -22\ -24]$) were activated by the main effect of $[+BN - BP + GN - GP]$. Abbreviations: B/G, child's bad/good attitude in the third phase; N/P, parent's negative/positive affective prosody in the fourth phase; A, auditory baseline (nonsense sound created by reversing a parent's utterance) in the fourth phase; MNI, Montréal Neurological Institute; BA, Brodmann's area.

Table 2
Activations related to affective prosody in sarcasm comprehension.

Cluster <i>p</i> -FWE	Cluster size	Z-value	MNI coordinates			Side	Location	BA
			x	y	z			
(a) Effect of Incongruity Change: contrast of [–BN +BP +GN –GP]								
0.012	314	4.31	–38	34	–8	L	IFG	47
(b) Effect of Prosodic Modulation: contrast of [+BN –BP +GN –GP]								
0.000	2689	6.63	2	28	40	R	ACC	32/8
0.000	2482	6.20	32	22	–12	R	AI	13
		5.31	52	18	10	R	IFG	44
		5.27	48	22	4	R	IFG	45
		4.46	50	38	–10	R	IFG	47
0.000	877	5.50	–30	18	–8	L	AI	13
0.001	497	5.03	0	–22	–24		Brainstem	

Abbreviations: IFG, inferior frontal gyrus; ACC, anterior cingulate cortex; AI, anterior insula; MNI, the Montréal Neurological Institute; BA, Brodmann's area; L, left; R, right.

regarding incongruity detection in sarcasm comprehension were excluded from the current study. For instance, while the current study was designed to investigate mechanisms of incongruity detection in situations where the affective prosody has a predicted role in modulating the semantic content of an utterance, there are potentially many other ways for the hearer to perceive an utterance as sarcastic. More specifically, previous studies suggested that recognition of incongruity between discourse context and emotional prosody (e.g., the combination of negative discourse context and positive emotional prosody) alone (e.g. Woodland and Voyer, 2011), or between the discourse context and semantic content of the utterance alone (e.g. a combination of positive discourse context and negative semantic content) (Colston, 2002), or even detection of incompatibility between the semantic content of the utterance and the emotional tone of voice alone (without context) (e.g. Bryant and Fox Tree, 2005), may be sufficient to allow the hearer to perceive an utterance as being sarcastic. In future research, the experimental paradigm used in this study would be effective not only in comparing and contrasting alternative mechanisms for detection of all potential sources of incongruity in sarcasm comprehension, but also for testing existing hypotheses about the function of the left IFG in utterance comprehension in general.

Acknowledgments

This study was partly supported by the Cooperative Study Program of National Institute for Physiological Sciences, Scientific Research on Innovative Areas grant #22101007 (to H.C.T.) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan (MEXT); by Challenging Exploratory Research grant #23650224 (to H.C.T.) from the Japan Society for the Promotion of Science (JSPS); and by Grants-in-Aid for Scientific Research (S) #21220005 (to N.S.), (A) #2124013 (to H.C.T., N.S.), (A) #15H01846 (to N.S.), and (B) #20330136 (to T.M., T.N.) from JSPS. Part of this study represents the results of the “Development of biomarker candidates for social behavior” and “Integrated research on neuropsychiatric disorders” projects carried out under the Strategic Research Program for Brain Science (MEXT).

References

- Ackerman, B.P., 1983. Form and function in children's understanding of ironic utterances. *J. Exp. Child. Psychol.* 35, 487–508.
- Banich, M.T., Milham, M.P., Jacobson, B.L., Webb, A., Wszalek, T., Cohen, N.J., Kramer, A.F., 2001. Attentional selection and the processing of task-irrelevant information: insights from fMRI examinations of the Stroop task. *Prog. Brain Res.* 134, 459–470.
- Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., Pike, B., 2000. Voice-selective areas in human auditory cortex. *Nature* 403 (6767), 309–312.
- Bryant, G.A., Fox Tree, J.E., 2005. Is there an ironic tone of voice? *Lang. Speech* 48 (3), 257–277.
- Capelli, C.A., Nakagawa, N., Madden, C.M., 1990. How children understand sarcasm: the role of context and intonation. *Child. Dev.* 61 (6), 1824–1841.
- Cheang, H.S., Pell, M.D., 2008. The sound of sarcasm. *Speech Commun.* 50 (5), 366–381.
- Colston, H., 2002. Pragmatic justifications for nonliteral gratitude acknowledgments: “Oh sure, anytime.”. *Metaphor Symb.* 17 (3), 205–226.
- Colston, H., O'Brien, J., 2000. Contrast and pragmatics in figurative language: Anything understatement can do, irony can do better. *J. Pragmat.* 32 (11), 1557–1583.
- Dapretto, M., Bookheimer, S.Y., 1999. Form and content: dissociating syntax and semantics in sentence comprehension. *Neuron* 24 (2), 427–432.
- Eickhoff, S.B., Paus, T., Caspers, S., Grosbras, M.H., Evans, A., Amunts, K., 2007. Assignment of functional activations to probabilistic cytoarchitectonic areas revisited. *NeuroImage* 36 (3), 511–521.
- Ethofer, T., Anders, S., Erb, M., Herbert, C., Wiethoff, S., Kissler, J., Grodd, W., Wildgruber, D., 2006. Cerebral pathways in processing of affective prosody: a dynamic causal modeling study. *NeuroImage* 30 (2), 580–587.
- Ethofer, T., Bretschner, J., Gschwind, M., Kreifelts, B., Wildgruber, D., Vuilleumier, P., 2012. Emotional voice areas: anatomic location, functional properties, and structural connections revealed by combined fMRI/DTI. *Cereb. Cortex* 22 (1), 191–200.
- Friston, K.J., Ashburner, J., Kiebel, S.J., Nichols, T.E., Penny, W.D., 2007. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Elsevier Academic Press, London.
- Frühholz, S., Grandjean, D., 2013. Processing of emotional vocalizations in bilateral inferior frontal cortex. *Neurosci. Biobehav. Rev.* 37 (10 Pt 2), 2847–2855.
- Gabrieli, J.D., Desmond, J.E., Demb, J.B., Wagner, A.D., Stone, M.V., Vaidya, C.J., Glover, G.H., 1996. Functional magnetic resonance imaging of semantic memory processes in the frontal lobes. *Psychol. Sci.* 7 (5), 278–283.
- Gerrig, R., Goldvarg, Y., 2000. Additive effects in the perception of sarcasm: situational disparity and echoic mention. *Metaphor Symb.* 15 (4), 197–208.
- Gibbs, R., 1986. On the psycholinguistics of sarcasm. *J. Exp. psychology: Gen.* 115 (1), 3–15.
- Giora, R., 1997. Understanding figurative and literal language: the graded salience hypothesis. *Cognit. Linguist.* 8 (3), 183–206.
- Hagoort, P., Hald, L., Bastiaansen, M., Petersson, K.M., 2004. Integration of word meaning and world knowledge in language comprehension. *Science* 304 (5669), 438–441.
- Hagoort, P., 2005. On Broca, brain, and binding: a new framework. *Trends Cognit. Sci.* 9 (9), 416–423.
- Ivanko, S.L., Pexman, P.M., 2003. Context incongruity and irony processing. *Discourse Process.* 35 (3), 241–279.
- Kapur, S., Rose, R., Liddle, P.F., Zipursky, R.B., Brown, G.M., Stuss, D., Houle, S., Tulving, E., 1994. The role of the left prefrontal cortex in verbal processing: semantic processing or willed action? *Neuroreport* 5 (16), 2193–2196.
- Katz, A.N., Lee, C.J., 1993. The role of authorial intent in determining verbal irony and metaphor. *Metaphor Symb. Act.* 8 (4), 257–279.
- Katz, A.N., Pexman, P.M., 1997. Interpreting figurative statements: speaker occupation can change metaphor to irony. *Metaphor Symb.* 12 (1), 19–41.
- Kreuz, R.J., Glucksberg, S., 1989. How to be sarcastic: the echoic reminder theory of verbal irony. *J. Exp. psychol.: Gen.* 118 (4), 374–386.
- Kumon-Nakamura, S., Glucksberg, S., Brown, M., 1995. How about another piece of pie: the allusional pretense theory of discourse irony. *J. Exp. psychology: Gen.* 124 (1), 3–21.
- Laval, V., Bert-Erboul, A., 2005. French-speaking children's understanding of sarcasm: the role of intonation and context. *J. Speech, Lang., Hear. Res.* 48 (3), 610–620.
- Leung, H., Hennessey, G., Drosopoulos, A., 2000. Signal detection using the radial

- basis function coupled map lattice. *IEEE Trans. Neural Netw. / A Publ. IEEE Neural Netw. Counc.* 11 (5), 1133–1151.
- Mead, L.A., Mayer, A.R., Bobholz, J.A., Woodley, S.J., Cunningham, J.M., Hammeke, T.A., Rao, S.M., 2002. Neural basis of the Stroop interference task: response competition or selective attention? *J. Int. Neuropsychol. Soc.* 8 (6), 735–742.
- Menenti, L., Petersson, K.M., Scheeringa, R., Hagoort, P., 2009. When elephants fly: differential sensitivity of right and left inferior frontal gyri to discourse and world knowledge. *J. Cognit. Neurosci.* 21 (12), 2358–2368.
- Menon, V., Uddin, L.Q., 2010. Saliency, switching, attention and control: a network model of insula function. *Brain Struct. Funct.* 214 (5–6), 655–667.
- Milham, M.P., Banich, M.T., Webb, A., Barad, V., Cohen, N.J., Wszalek, T., Kramer, A.F., 2001. The relative involvement of anterior cingulate and prefrontal cortex in attentional control depends on nature of conflict. *Brain Res.: Cognit. Brain Res.* 12 (3), 467–473.
- Mitchell, R.L., 2006. How does the brain mediate interpretation of incongruent auditory emotions? The neural response to prosody in the presence of conflicting lexico-semantic cues. *Eur. J. Neurosci.* 24 (12), 3611–3618.
- Mitchell, R.L., Ross, E.D., 2013. Attitudinal prosody: what we know and directions for future study. *Neurosci. Biobehav. Rev.* 37 (3), 471–479.
- Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9 (1), 97–113.
- Pell, M.D., 2006. Judging emotion and attitudes from prosody following brain damage. *Prog. Brain Res.* 156, 303–317.
- Petrides, M., 2005. Lateral prefrontal cortex: architectonic and functional organization. *Philos. Trans. R. Soc. London: Ser. B, Biol. Sci.* 360 (1456), 781–795.
- Peterson, B.S., Kane, M.J., Alexander, G.M., Lacadie, C., Skudlarski, P., Leung, H.C., May, J., Gore, J.C., 2002. An event-related functional MRI study comparing interference effects in the Simon and Stroop tasks. *Brain Res.: Cognit. Brain Res.* 13 (3), 427–440.
- Pexman, P., 2008. It's fascinating research: the cognition of verbal irony. *Curr. Dir. Psychol. Sci.* 17 (4), 286–290.
- Rapp, A.M., Leube, D.T., Erb, M., Grodd, W., Kircher, T.T., 2004. Neural correlates of metaphor processing. *Brain Res.: Cognit. Brain Res.* 20 (3), 395–402.
- Rapp, A.M., Mutschler, D.E., Wild, B., Erb, M., Lengsfeld, I., Saur, R., Grodd, W., 2010. Neural correlates of irony comprehension: the role of schizotypal personality traits. *Brain Lang.* 113 (1), 1–12.
- Rapp, A.M., Erb, M., Grodd, W., Bartels, M., Markert, K., 2011. Neural correlates of metonymy resolution. *Brain Lang.* 119 (3), 196–205.
- Rockwell, P., 2007. Vocal features of conversational sarcasm: a comparison of methods. *J. Psycholinguist. Res.* 36 (5), 361–369.
- Ross, E.D., 2000. Affective prosody and the aprosodias. In: Mesulam, M.-M. (Ed.), *Principles of Behavioral and Cognitive Neurology*. Oxford University Press, New York, pp. 316–331.
- Sakai, K.L., 2005. Language acquisition and brain development. *Science* 310 (5749), 815–819.
- Schirmer, A., Kotz, S.A., 2006. Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends Cognit. Sci.* 10 (1), 24–30.
- Schirmer, A., Zysset, S., Kotz, S.A., Yves von Cramon, D., 2004. Gender differences in the activation of inferior frontal cortex during emotional speech perception. *NeuroImage* 21 (3), 1114–1123.
- Shamay-Tsoory, S.G., Tomer, R., Aharon-Peretz, J., 2005. The neuroanatomical basis of understanding sarcasm and its relationship to social cognition. *Neuropsychologia* 19 (3), 288–300.
- Spotorno, N., Koun, E., Prado, J., Van Der Henst, J.B., Noveck, I.A., 2012. Neural evidence that utterance-processing entails mentalizing: the case of irony. *NeuroImage* 63 (1), 25–39.
- 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. *Proc. Natl. Acad. Sci. USA* 105, 12569–12574.
- Thompson-Schill, S.L., D'Esposito, M., Aguirre, G.K., Farah, M.J., 1997. Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. *Proc. Natl. Acad. Sci. USA* 94 (26), 14792–14797.
- Thompson-Schill, S.L., D'Esposito, M., Kan, I.P., 1999. Effects of repetition and competition on activity in left prefrontal cortex during word generation. *Neuron* 23 (3), 513–522.
- Thompson-Schill, S.L., 2003. Neuroimaging studies of semantic memory: inferring "how" from "where". *Neuropsychologia* 41 (3), 280–292.
- Turken, A.U., Dronkers, N.F., 2011. The neural architecture of the language comprehension network: converging evidence from lesion and connectivity analyses. *Front. Syst. Neurosci.* 5 (1), 1–20.
- Uchiyama, H., Seki, A., Kageyama, H., Saito, D.N., Koeda, T., Ohno, K., Sadato, N., 2006. Neural substrates of sarcasm: a functional magnetic-resonance imaging study. *Brain Res.* 1124 (1), 100–110.
- Utsumi, A., 2000. Verbal irony as implicit display of ironic environment: distinguishing ironic utterances from nonirony. *J. Pragmat.* 32, 1777–1806.
- Wagner, A.D., Desmond, J.E., Demb, J.B., Glover, G.H., Gabrieli, J.D., 1997. Semantic repetition priming for verbal and pictorial knowledge: a functional MRI study of left inferior prefrontal cortex. *J. Cognit. Neurosci.* 9 (6), 714–726.
- Wang, A.T., Lee, S.S., Sigman, M., Dapretto, M., 2006a. Neural basis of irony comprehension in children with autism: the role of prosody and context. *Brain* 129 (4), 932–943.
- Wang, A.T., Lee, S.S., Sigman, M., Dapretto, M., 2006b. Developmental changes in the neural basis of interpreting communicative intent. *Soc. Cognit. Affect. Neurosci.* 1 (2), 107–121.
- Wildgruber, D., Ackermann, H., Kreifelts, B., Ethofer, T., 2006. Cerebral processing of linguistic and emotional prosody: fMRI studies. *Prog. Brain Res.* 156, 249–268.
- Wildgruber, D., Ethofer, T., Grandjean, D., Kreifelts, B., 2009. A cerebral network model of speech prosody comprehension. *Int. J. Speech-Lang. Pathol.* 11 (4), 277–281.
- Wilson, D., Sperber, D., 2012. *Meaning and Relevance*. Cambridge University Press, Cambridge.
- Wittfoth, M., Schroder, C., Schardt, D.M., Dengler, R., Heinze, H.J., Kotz, S.A., 2010. On emotional conflict: interference resolution of happy and angry prosody reveals valence-specific effects. *Cereb. Cortex* 20 (2), 383–392.
- Woodland, J., Voyer, D., 2011. Context and intonation in the perception of sarcasm. *Metaphor Symb.* 26 (3), 227–239.
- Zemleni, M.Z., Haverkort, M., Renken, R., A Stowe, L., 2007. Evidence for bilateral involvement in idiom comprehension: an fMRI study. *NeuroImage* 34 (3), 1280–1291.
- Zysset, S., Muller, K., Lohmann, G., Von Cramon, D.Y., 2001. Color-word matching stroop task: separating interference and response conflict. *NeuroImage* 13 (1), 29–36.