Journal of Neurolinguistics 42 (2017) 49–62

Contents lists available at ScienceDirect

Journal of Neurolinguistics

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

A re-visit of three-stage humor processing with readers' surprise, comprehension, and funniness ratings: An ERP study

Li-Chuan Ku [a], Yen-Ju Feng [a], Yu-Chen Chan [b], Ching-Lin Wu [a], Hsueh-Chih Chen [a,*]

[a] Department of Educational Psychology and Counseling, National Taiwan Normal University, Taipei, Taiwan
[b] Institute of Learning Sciences, National Tsing Hua University, Hsinchu, Taiwan

ARTICLE INFO

Article history:
Received 31 December 2015
Received in revised form 7 November 2016
Accepted 14 November 2016
Available online 29 November 2016

Keywords:
Surprise
Comprehension
Amusement
Humor
Event-related potential (ERP)

ABSTRACT

Humor processing can be divided into three sub-stages including incongruity detection, incongruity resolution, and elaboration (Chan et al., 2012, 2013; Feng et al., 2014). However, few studies have investigated the three-stage model of humor processing with readers' surprise, comprehensibility and funniness levels, and little discussion has been devoted to its biological underpinning. To verify the credibility of the three-stage model, electroencephalography (EEG) was utilized in corroboration with two types of stimuli including jokes and non-jokes in the present research. Participants were categorized into high vs. low score groups based on their rating scores of surprise, comprehension, and funniness to joke stimuli. The between-group analyses showed that compared with the less surprised group, highly surprised people elicited a primarily larger N400, which may suggest more incongruity perceived in reading jokes. Additionally, good comprehenders mainly elicited a larger P600, probably indicating a more successful resolution of detected incongruity in comparison with poor comprehenders. Finally, the highly amused group elicited a larger late positive potential (LPP) compared with the less amused group, which could reflect more affective elaboration of jokes. Participants' surprise, comprehension, and funniness levels had smaller impacts on other chief electrophysiological components, with the effects varying with different group contrasts. These results provided the evidence that different degrees of surprise, comprehensibility, and amusement to jokes would influence the three sub-stages (incongruity detection, incongruity resolution, and elaboration) respectively in humor processing. The current study thus generally re-verified the stability of the three-stage model through participants' behavioral ratings which had seldom been touched upon.

1. Introduction

Humor is a unique ability in human beings. Successful humor processing is really complicated because it involves a cognitive juxtaposition of mental sets, followed by an affective feeling of amusement. Suls' (1972) incongruity resolution theory was the earliest to propose that people will make a prediction about what the ending, or a punch line, will be when

* Corresponding author. Department of Educational Psychology and Counseling, National Taiwan Normal University, No. 162, Sec. 1, Heping E. Rd., Da-an District, Taipei City 10610, Taiwan.
E-mail address: chcjyh@ntnu.edu.tw (H.-C. Chen).

http://dx.doi.org/10.1016/j.jneuroling.2016.11.008
0911-6044/

reading a text. Once the actual punch line is different from the predicted one, surprise occurs to motivate one to search for possible logical rules and to incorporate the new information that causes initial incongruity. Successful resolution of incongruity could thus lead to a sense of enjoyment. Wyer and Collins' (1992) comprehension and elaboration theory later attempted to break down the stream of humor processing into the comprehension stage including the detection of incongruity and its resolution, followed by the elaboration process to form a mirth experience. In the elaboration process, further inferences are made based on implicit or newly-understood information, and it can lead to the appraisals of humorous stimuli.

Following Wyer and Collins' (1992) theory, Chan, Chou, Chen, and Liang's (2012) functional magnetic resonance imaging (fMRI) study successfully provided the neural correlates of different cognitive and affective components in humor processing. By comparing funny (including stages of humor comprehension and elaboration), unfunny (neither humor comprehension nor elaboration needed), and garden path sentences (including only the comprehension stage), their results showed that the stage of humor comprehension (containing both incongruity detection and resolution) was associated with enhanced activities in the bilateral inferior frontal gyri (IFG) and left superior frontal gyrus (SFG), while the elaboration process of humor was correlated to the activities in left ventromedial prefrontal cortex (VMPFC), bilateral amygdalae, and parahippocampal gyri.

Also, by manipulating funny (including incongruity detection and resolution), unfunny (neither incongruity detection nor resolution needed), and nonsensical sentences (including only incongruity detection), Chan et al. (2013) had further dissociated the neural circuits underlying incongruity detection and resolution in humor comprehension. Incongruity detection was found to be associated with greater neural activations in the right middle frontal and right temporal gyrus (rMFG, rMTG), while the resolution of incongruity activated the left SFG and left inferior parietal lobes (IPL) more. Taken together, these fMRI studies had integrated the traditional two-stage model of humor processing into a three-stage one, namely the incongruity detection, resolution, and elaboration.

However, these cognitive theories of humor processing often put different emphasis on surprise, comprehension, and the appraisal of humor, the three critical components in humor processing. According to recent fMRI studies (Kohn, Kellermann, Gur, Schneider, & Habel, 2011; Mobbs, Hagan, Azim, Menon, & Reiss, 2005), there exists great differences in processing of verbal humor among different readers. Therefore, even though Chan et al.'s (2012, 2013) studies had proposed a three-stage model of humor processing, between-group variances in the perception of surprise, readers' comprehensibility, or emotion arousal (amusement) during humor processing were not taken into consideration, given that humorous forms of occurrence, reception and presentation can be varied greatly across individuals, cultures and genders (Alden, Hoyer, & Lee, 1993; Hehl & Ruch, 1985). More importantly, due to limited temporal resolutions of fMRI techniques, a closer look into the dynamic and the processing stream in each sub-stage of humor processing still seemed unattainable.

Coulson and Kutas's (2001) research is one of the few ERP studies that examined readers' differences in comprehensibility during humor processing. By utilizing one-liner jokes along with non-jokes, the authors investigated neural mechanisms underlying the stages of incongruity detection and resolution by grouping participants into good and poor comprehenders based on their response to comprehension questions. The results indicated that compared with poor comprehenders, good comprehenders exhibited more posterior N400 (350–500 ms) amplitudes and an enhanced late posterior positivity (500–900 ms) when reading high-constraint jokes. Both the high-/low-constraint jokes elicited a left-lateralized sustained negativity (500–900 ms) in the two groups. The authors further proposed that the observed N400 effect could signal the increased difficulty of contextual integration and the brain's sensibility to the inconsistent expectations based on the frames retrieved from long-term memory when reading punch lines. Additionally, the late positivity may comprise a broad P3 family that could reflect the violation of frame-level expectations and/or an orienting reaction. Finally, the sustained anterior negativity may be linked to enhanced working memory load to resolve incongruity during frame-shifting processes in joke comprehension.

Although Coulson and Kutas' (2001) study has provided important information regarding the influences of comprehensibility in the processing of humor, they still failed to map ERP effects to the surprise and resolution stages of incongruity in joke comprehension specifically. More importantly, they could not exclude the possibility that the results they found could be influenced by participants' different degrees of perceived surprise and funniness to the jokes.

More recently, with funny, congruous (unfunny) and incongruous (nonsensical) question-answer type sentences, Marinkovic et al. (2011) found that funny punch lines generated the smallest N400m (350–500 ms) in left anteroventral temporal lobe, which may imply initial lexical-semantic analysis, by recording ERPs and magnetoencephalography (MEG) simultaneously. In addition, these funny sentences elicited the largest posterior P600 (700–1150 ms) with increased activation in anterior medial prefrontal cortex and right dorsolateral prefrontal cortex, which was suggested to reflect ambiguity detection and the subsequent interpretative-integrative processing. On the contrary, with discourse coherent, incoherent, and joke endings, Mayerhofer and Schacht's (2015) failed to find the late left anterior negativity (500–700 ms) or posterior P600s (700–1000 ms) as seen in the previous studies (Coulson & Kutas, 2001; Marinkovic et al., 2011). Joke endings elicited larger N400 (250–500 ms) amplitudes and the late frontal positivity (700–1000 ms) than coherent endings. The authors argued that N400 could reflect expectation violation of readers' initial interpretation at punch lines, while the late frontal positivity probably indexed further emotion responses, which was supported by the data of participants' larger pupil diameters when reading joke endings.

To clearly dissociate the overlapped mapping of cognitive processes with electrophysiological activities underlying humor processing, Feng, Chan, and Chen (2014) investigated the temporal dynamics of cortical activations with funny, unfunny, and nonsensical question-answer type sentences. The within-subject comparisons indicated that nonsensical sentences elicited the largest negative event-related potential (ERP) deflection (N400) between 350 and 500 ms (ms). Also, jokes elicited greater N400 amplitudes than non-joke sentences did at fronto-central scalp regions, possibly reflecting incongruity detection in humor. Between 500 and 700 ms, jokes as well as non-jokes elicited a more positive deflection (P600) at central-parietal scalp regions than nonsensical sentences did, possibly indicating a reanalysis process during incongruity resolution. Moreover, jokes elicited the most positive slow-wave activity between 800 and 1500 ms at central-parietal scalp regions compared with non-jokes and nonsensical sentences, which might be related to emotional processing during the elaboration process of humor processing. Based on these results, the authors concluded that the cognitive functions involved in humor processing can be indexed by N400, P600 and late positive potentials (LPP) respectively.

However, in these ERP studies, although the stimuli across different experimental conditions were delicately controlled for surprise/expectancy (i.e., whether the ending was expected in the context), plausibility/comprehensibility (i.e., whether the ending made sense in the context), and humorous potential (i.e., whether the ending was funny in the context), little has been focused on the possible impact of readers' variances in humor processing per se from a better temporal resolutions. More importantly, it was argued that individual comprehension ability of the jokes, contextual influences (related to the expectation built from the joke setups), or weak effect sizes due to the limited number of participants, could all render the inconsistent ERP results (Mayerhofer & Schacht, 2015). Thus, a detailed investigation of participants' subjective evaluation towards humorous stimuli would be necessary to re-verify the three-stage model of humor processing.

Consequently, the goal of the present study is not only to examine whether N400, P600 and LPP would be influenced by readers' surprise, comprehension, and funniness ratings to jokes, but also to incorporate the role of these behavioral patterns into the three-stage model of humor processing with corresponding neurophysiological activities. Adopting Feng et al.'s (2014) experimental stimuli, their question/answer (setups/punch lines) type statements were utilized in the present study as well to time-lock the occurrence of humor more precisely. Setup sentences in jokes were collected to induce incongruity detection, incongruity resolution and the further elaboration process of humor, while setup sentences in non-jokes were logically matched for corresponding punch lines to serve as the baseline.

According to the previous research (Coulson & Kutas, 2001; Du et al., 2013; Feng et al., 2014; Hamm, Johnson, & Kirk, 2002; Mayerhofer & Schacht, 2015), N400 is sensitive to semantic violations and expectation, often reflecting the detection of incongruity and/or expectation violation with a posterior (sometimes anterior) scalp distribution in humor processing. We thus hypothesized that people with high surprise will elicit larger N400 (350–500 ms) amplitudes than those with low surprise due to a stronger violation of expectation at punch lines. In addition, P600 (500–700 ms) may reflect syntactic violations and sentence re-analysis, with a centro-parietal scalp distribution (Coulson, King, & Kutas, 1998; Donchin, 1981; Hahne & Friederici, 1999). In joke processing, P600 was often associated with the updating/integrating of novel associations or interpretation in the resolution of incongruity (Du et al., 2013; Feng et al., 2014; Marinkovic et al., 2011). Hence, we expected that good comprehenders would induce larger amplitudes of P600 than poor comprehenders, as found in Coulson and Kutas's (2001) study. Lastly, we predicted that if the sustained LPP can reflect the elaboration of amusement in humor processing (affective processing), then the highly amused people would induce stronger LPPs than the less amused ones at both pre-frontal and central-parietal regions, according to Moratti, Saugar, and Strange (2011).

2. Methods

2.1. Participants

Thirty-two native Chinese adults (19 females), aged between 20 and 32 year (Mean ± SD: 23.03 ± 2.83 years), took part in the experiment. They were all right-handed, free of hearing impairments, had normal or corrected-to-normal vision, and had no neurological disorders or psychiatric illness. All the participants were naïve to the experimental tasks. They received verbal and written instructions regarding all the details of the experiment before the formal experiment to make sure they were familiar with the experimental procedures. They were also required to provide the informed consent according to the guidelines approved by the Research Ethics Office of National Taiwan University (Taipei, Taiwan), and were allowed to withdraw at any time. Each participant received a sum of NT. 500 dollars for participating in the experiment as compensation for the inconvenience and commuting to the experimental sites, with the amount approved by the same local ethical committee.

2.2. Experimental design

The present study adopted a mixed design approach. Stimulus Type (joke, non-joke) and Electrode (Fz, FCz, Cz, CPz Pz) were manipulated to serve as within-subject independent variables. In addition, all the participants went through three times of classifications based on their ratings of surprise, comprehension, and funniness to the joke stimuli. The mean scores of surprise, comprehension, and funniness ratings to all the joke stimuli for each participant were calculated first. Accordingly, the participants were categorized into high vs. low score groups based on the medians of these mean scores. Firstly, participants were divided into a highly and a less surprised group based on their median rating scores of surprise (Median = 2.68). Second, participants were divided into a good and a poor comprehension group based on their median rating scores of comprehensibility (Median = 3.24). Finally, participants were divided into a highly and a less amused group based on their median rating scores of funniness (Median = 2.67). The number of people were identical (N = 16) in all the subgroups. Group (high vs. low score group) further served as a between-subject factor in the comparison of ERP mean amplitudes elicited by jokes in the present study. The measurement of ERP mean amplitudes served as the dependent variable in both the omnibus and between-group analysis.

2.3. Material

To time-lock neural activities precisely, the presentation of the setup sentences was carried out by a question followed by a corresponding concise answer that served as a punch line. First, 30 funny question/answer type statements were collected from Feng et al. (2014)'s pool of Chinese riddles to be used as jokes in the current study. All the selected jokes were semantically incongruent riddles to avoid confounding of the other types of verbal humor (e.g. phonological puns, syntactic jokes, or pragmatic jokes). A question/answer type statement for the joke, for example, would be "睡美人這輩子最怕遇到什麼事？失眠" (What is the Sleeping Beauty most afraid of? Insomnia). The critical answer is surprising but fits in a clever and unexpected way. Participants found this joke funny after realizing that the Sleeping Beauty is unlikely to get "失眠" ("insomnia"), concerning their previous knowledge of this fairy tale. Subsequently, 30 coherent setups with the corresponding punch lines were randomly selected from the same pool to serve as non-jokes and baselines in the experiment (e.g. "大地震後沿海通常需要防範什麼？海嘯" (What should the precaution be taken against in coastal areas after big earthquakes? Tsunami.)). In total, there were 60 pairs of sentences included in the present study, consisting of two types of stimuli: jokes and non-jokes.

Prior to the formal ERP experiment, all of the experimental sentences had been validated by another group of participants with nine-point Likert scales on the ratings of surprise, comprehension, and funniness (Fig. 1). The statistical results showed that for both surprise and funniness ratings, jokes (Surprise: Mean = 6.27, SD = 0.71; Funniness: Mean = 6.76, SD = 0.37) scored higher than non-jokes (Surprise: Mean = 3.61, SD = 0.74; Funniness: Mean = 3.75, SD = 0.40) ($ps < 0.001$). However, there was no difference in comprehension ratings between jokes (Mean = 8.05, SD = 0.41) and non-jokes (Mean = 8.06, SD = 0.62) ($p = 0.981$). Afterwards, we randomly divided the 60 sets of experimental stimuli into two blocks. The order of presentation of each stimulus type was randomized. The lengths of the setup sentences were controlled across the two conditions and confined to 11–14 Chinese characters. To avoid excessive eye movements, the lengths of the critical punch lines were also controlled and strictly constrained to be within 2–4 Chinese characters.

2.4. Procedure

The experiment was carried out in Educational Neuroscience Lab at National Taiwan Normal University. Participants were seated approximately 110 cm from the computer screen (Visual angle: 1.8–3.1°) in a sound-proofing room. All the stimuli were presented centrally on the screen, designed by E-Prime software (Psychology Software Tools, Inc.). During the formal experiment, participants were instructed to read experimental sentences with correspondent punch lines and, following each set, to rate on a 4-point Likert scale: (a) surprise: how surprised they feel at the punch line after reading the setup sentence (not surprised at all/not surprised/surprised/extremely surprised) (b) comprehensibility: how well they understood the punch line's meaning considering the corresponding setup (did not comprehend at all/did not comprehend/comprehended/comprehended extremely well), and (c) funniness: how funny they thought the punch line was (not funny at all/not funny/funny/extremely funny). Each experimental trial started with a central fixation mark lasting for 1000 ms to orient the participant to the center of the screen. Then, a setup sentence was presented and remained on the screen for 4000 ms, followed by a punch line for 2000 ms. To avoid eye movement artifacts elicited by reading the stimuli, a fixation mark also appeared between the setup sentence and the punch line for 4000 ms. An illustration of experimental procedure was shown in Fig. 2.

2.5. EEG acquisition

Electroencephalogram (EEG) data were continuously recorded from 32 scalp electrodes that were mounted on an elastic cap according to the international 10–20 system. All electrodes were online referenced to the average of the left (A1) and right (A2) mastoids. Eye blinks and vertical eye movements were monitored using electrodes located above and below the left eye. Horizontal electro-oculogram (EOG) data were recorded from two electrodes placed 1.5 cm lateral to the external canthi of the left and right eye. A ground electrode was placed on the forehead. The continuous EEG signal was digitized at a sampling rate of 500 Hz, and amplified with a bandpass from 0.05 to 70 Hz by using the SynAmps2 (Neuroscan, Inc.) amplifier. The electrode/skin impedance was kept below 5 kΩ during recording.

2.6. Data analysis

EEG recordings were processed and analyzed offline with Scan 4.5 software (Compumedics Ltd., Australia). The raw EEG data then went through the ocular artifact correction algorithm (ARTCOR) implemented in Edit 4.5 (Neuroscan, 2003) to remove possible blinks. The ERP waveforms were time-locked to the onset of the punch line and epoched by setting the interval as 100 ms before and 2000 ms after the stimulus onset, with the pre-stimulus interval of −100 to 0 ms as the baseline.

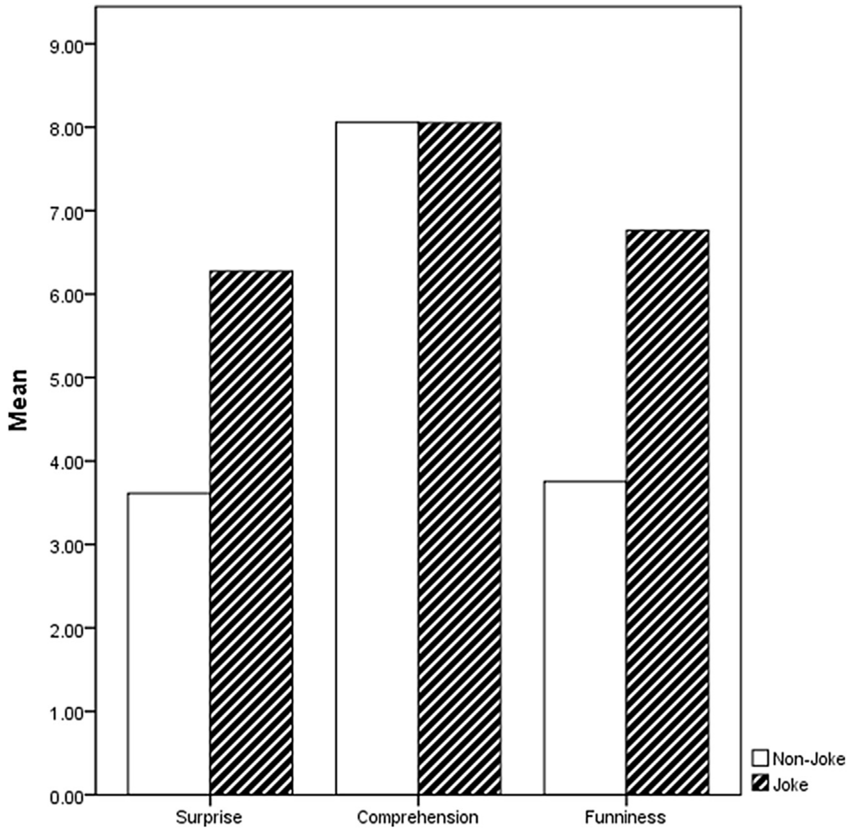

Fig. 1. In the ratings of surprise and funniness, jokes scored higher than non-jokes. In the rating of comprehensibility, there was no difference between jokes and non-jokes.

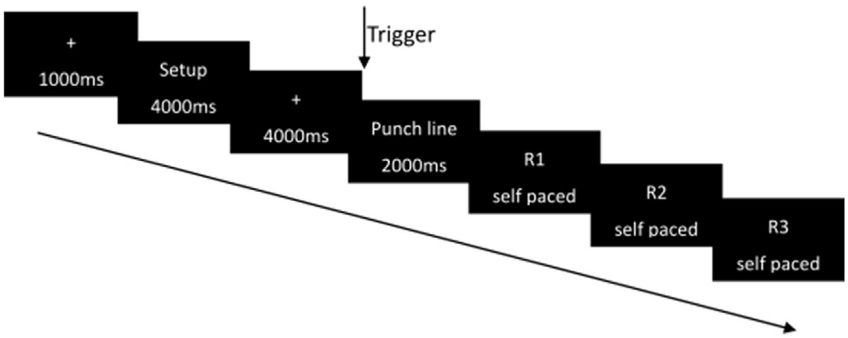

Fig. 2. Experimental procedure adopted from Feng et al. (2014). The ERP recording was time locked to the onset of punch lines. After each experimental trial, participants were instructed to rate on a 4-point Likert scale (R1) surprise, (R2) comprehensibility, and (R3) funniness.

All epochs were band-pass filtered with frequency values set as 0.1–30 Hz using digital, zero-phase shift filtering. Finally, trials contaminated with artifacts due to amplifier clipping, burst of electromyography activity, or peak to peak deflection exceeding ± 100 μV were excluded from following sorting and averaging. Trials in which participants rated "not funny at all" for the joke stimuli, and "extremely funny" for the non-joke stimuli were also excluded from the following analysis.

To re-verify the three-stage model of humor processing in the previous research (Feng et al., 2014), an omnibus statistical analysis was first conducted on the ERP mean amplitudes in time windows of 350–500 ms (incongruity detection), 500–700 ms (incongruity resolution) and 700–1000 ms (affective processing) post-stimulus. For each time interval, ERP effects were assessed by performing a repeated-measure analysis of variance (ANOVA) (Greenhouse & Geisser, 1959) with the factors of Stimulus Type (joke, non-joke) and Electrode (Fz, FCz, Cz, CPz, Pz).

In addition, to investigate participants' variances in reading jokes, mixed design ANOVAs with the between-subject factor of Group (high vs. low score group) and the within-subject factor of Electrode (Fz, FCz, Cz, CPz, Pz) were performed in the same time windows of N400 (350–500 ms), P600 (500–700 ms), and LPP (700–1000 ms), and an additional time window of P2 (200–250 ms). For each time interval, three mixed design ANOVAs were conducted for groups with high vs. low surprise, groups with high vs. low comprehension, and groups with high vs. low funniness ratings separately. An alpha level of 0.05 was used for all statistical tests. All pairwise post-hoc analyses were Bonferroni corrected. To adjust the degrees of freedom (DF) of the F-values for violations of the sphericity assumption, Greenhouse-Geisser corrections were applied.

3. Results

3.1. Behavioral results

The summary of rating scores and reaction times (RTs) were illustrated in Table 1. Pair t-tests were performed on rating scores and RTs between jokes and non-jokes for surprise, comprehension, and funniness ratings respectively. The results showed that in consistent with a priori validation data, jokes scored higher in both surprise and funniness ratings than non-jokes ($t(31) = 12.706$, $p < 0.001$; $t(31) = 12.565$, $p < 0.001$), while there is no difference in the rating scores of comprehension ($t(31) = -1.354$, $p = 0.185$). There is no differences of RTs for the three ratings between the two types of stimuli (Surprise: $t(31) = 1.333$, $p = 0.192$; Comprehension: $t(31) = 0.348$, $p = 0.730$; Funniness: $t(31) = -0.949$, $p = 0.350$), as the responses were collected from the onset of rating cues, rather than the punch lines. A further correlation analysis showed that participants' funniness scores were positively correlated with their comprehension scores ($r = 0.551$, $p = 0.001$), but not with surprise scores ($r = 0.118$, $p = 0.521$).

In order to elucidate the between-group differences on rating scores, as shown in Table 2, independent t-tests between high vs. low score groups were also conducted on the three rating scores to jokes respectively. To our expectation, the highly surprised group scored higher than the less surprised group in surprise ratings ($t(30) = -6.455$, $p < 0.001$), but not in comprehension and funniness ratings ($t(30) = 1.021$, $p = 0.315$; $t(30) = -0.213$, $p = 0.833$). The good comprehension group also scored higher on ratings of comprehension and funniness than the poor comprehension group ($t(30) = -8.100$, $p < 0.001$; $t(30) = -3.595$, $p = 0.001$), but not on ratings of surprise ($t(30) = 0.659$, $p = 0.515$). Finally, the highly amused group scored higher on ratings of funniness and comprehension than the less amused group ($t(30) = -6.167$, $p < 0.001$; $t(30) = -2.218$, $p = 0.034$), but not on ratings of surprise ($t(30) = -0.243$, $p = 0.810$).

Furthermore, the mean ratios of expected categorization of jokes as humorous stimuli (scored 3 and 4 on funniness ratings) and non-jokes as non-humorous stimuli (scored 1 and 2 on funniness ratings) are shown in Table 3. Overall, this ratio for non-jokes was higher than that of jokes ($t(31) = -6.538$, $p < 0.001$). Three two-way repeated-measure ANOVAs with the between-subject factor of Group (high vs. low score group) and the within-subject factor of Stimulus Type (joke, non-joke) were performed on the collapsed data. For the group contrast of surprise, the results revealed only a main effect of Stimulus Type ($F(1, 30) = 41.603$, $p < 0.001$, $\eta_p^2 = 0.581$). Generally, non-jokes (Mean = 94.0%, SE = 1.5%) had more anticipated responses than jokes (Mean = 61.5%, SE = 4.2%). No other main effect or interaction was found (Group: $F(1, 30) = 0.341$, $p = 0.561$, $\eta_p^2 = 0.011$; Group X Stimulus Type: $F(1, 30) = 0.173$, $p = 0.680$, $\eta_p^2 = 0.006$). In addition, the good vs. poor comprehension group showed main effects of Stimulus Type ($F(1, 30) = 51.655$, $p < 0.001$, $\eta_p^2 = 0.633$) and Group ($F(1, 30) = 9.648$, $p = 0.004$, $\eta_p^2 = 0.243$), and an interaction of the two factors ($F(1, 30) = 7.464$, $p = 0.010$, $\eta_p^2 = 0.199$). Follow-up analysis revealed that the good comprehension group had more correct categorization than the poor comprehension group only in funniness ratings of jokes ($t(30) = -3.087$, $p = 0.004$). In highly vs. less amused group, both the main effects of Group ($F(1, 30) = 22.250$, $p < 0.001$, $\eta_p^2 = 0.426$) and Stimulus Type ($F(1, 30) = 89.336$, $p < 0.001$, $\eta_p^2 = 0.749$), and the interaction between the two factors ($F(1, 30) = 34.792$, $p < 0.001$, $\eta_p^2 = 0.537$) were significant. Further analysis showed that the highly amused group had more expected responses than the less amused group did in funniness ratings of jokes ($t(30) = -5.946$, $p < 0.001$), while the reverse pattern was seen in those of non-jokes ($t(30) = 2.476$, $p = 0.023$).

Table 1
The mean and standard deviation (SD) of rating scores and reaction time (RT).

Rating score	Joke		Non-joke	
	Mean scores (SD)	Mean RT (SD)	Mean scores (SD)	Mean RT (SD)
Surprise	2.69 (0.42)	1292.79 (263.37)	1.71 (0.33)	1246.71 (255.49)
Comprehension	3.36 (0.40)	1103.16 (239.76)	3.41 (0.33)	1092.72 (234.61)
Funniness	2.79 (0.41)	1055.07 (306.50)	1.67 (0.30)	1088.37 (277.77)

Table 2
The rating scores on jokes collapsed by highly vs. less surprised, comprehended, and amused groups.

Rating score	Surprise		Comprehension		Funniness	
	Mean	SD	Mean	SD	Mean	SD
Highly surprised group	3.00	0.29	3.29	0.39	2.80	0.42
Less surprised group	2.37	0.27	3.43	0.40	2.77	0.41
Good comprehension group	2.64	0.42	3.69	0.25	3.00	0.36
Poor comprehension group	2.74	0.44	3.04	0.19	2.56	0.34
Highly amused group	2.71	0.45	3.51	0.38	3.09	0.32
Less amused group	2.67	0.41	3.22	0.36	2.49	0.23

3.2. ERP results

3.2.1. Omnibus analysis

ERPs elicited by jokes and non-jokes are shown in Fig. 3. As the typical ERPs to visually presented words, all stimuli contained the clear N1 and P2 complex. By visual inspection, the mean amplitudes of the two stimulus types showed clear deflections in the time windows of 350–1000 ms post-stimulus. Below are the results of the analysis of ERPs measured in the time interval of N400 (350–500 ms post-stimulus) and the late positivity complex (500–700 ms and 700–1000 ms post-stimulus). In each time window, a two-way repeated-measure ANOVA with the within-subject factors of Stimulus Type (joke, non-joke) and Electrode (Fz, FCz, Cz, CPz, Pz) was conducted accordingly.

3.2.1.1. 350–500 ms. The statistical results showed a main effect of Electrode ($F(4, 124) = 8.780$, $p < 0.001$, $\eta_p^2 = 0.221$), and a marginal effect of Stimulus type ($F(1, 31) = 3.748$, $p = 0.062$, $\eta_p^2 = 0.108$). There was no interaction between Stimulus Type and Electrode ($F(4, 124) = 1.080$, $p = 0.342$, $\eta_p^2 = 0.034$). Further comparison showed that anterior sites (Fz and FCz) elicited larger negativity than posterior sites (CPz and Pz). Jokes tended to elicit slightly larger negativity (Mean = 7.838, SE = 0.820) than non-jokes (Mean = 8.920, SE = 0.950).

3.2.1.2. 500–700 ms. The two-way repeated-measure ANOVA revealed only a main effect of Electrode ($F(4, 124) = 15.783$, $p < 0.001$, $\eta_p^2 = 0.337$), but not of Stimulus Type ($F(1, 31) = 2.975$, $p = 0.095$, $\eta_p^2 = 0.088$). There was no interaction between Stimulus Type and Electrode ($F(4, 124) = 1.997$, $p = 0.141$, $\eta_p^2 = 0.061$). Follow-up analysis showed that the observed positivity elicited at posterior site (Cz, CPz and Pz) was larger than anterior sites (Fz and FCz).

3.2.1.3. 700–1000 ms. The analysis indicated an interaction of Stimulus Type and Electrode ($F(4, 124) = 3.961$, $p = 0.018$, $\eta_p^2 = 0.113$), as well as main effects of Stimulus Type ($F(1, 31) = 25.529$, $p < 0.001$, $\eta_p^2 = 0.452$) and Electrode ($F(4, 124) = 6.368$, $p = 0.002$, $\eta_p^2 = 0.170$). Further analysis of simple main effects revealed a larger positivity elicited by jokes compared with non-jokes at each of the midline electrodes (Fz: $t(31) = -5.713$, $p < 0.001$; FCz: $t(31) = -5.947$, $p < 0.001$; Cz: $t(31) = -3.903$, $p < 0.001$; CPz: $t(31) = -3.671$, $p = 0.001$; Pz: $t(31) = -2.655$, $p = 0.012$).

3.2.2. Between-group analysis

By visual inspection of Figs. 4–6, ERPs elicited by jokes in each group contrast based on the three types of participant classification indicated different deflections in both early (P2, N400) and later (P600, LPP) time windows. Therefore, separated mixed-design ANOVAs were conducted with the between-subject factor of Group (high v.s. low score groups) and the within-subject factor of Electrode (Fz, FCz, Cz, CPz, Pz) in the time windows of 200–250 ms, 350–500 ms, 500–700 ms, and 700–1000 ms.

3.2.2.1. Highly vs. less surprised group. To examine whether the N400 component would be affected by participants' surprise feelings to the joke stimuli, a two-way mixed-design ANOVA involving the between-subject factor of Group (highly/less surprised group) and the within-subject factor of Electrode (Fz, FCz, Cz, CPz, Pz) was conducted in the time window of 350 and 500 ms. The statistical results revealed main effects of Group ($F(1, 30) = 5.351$, $p = 0.028$, $\eta_p^2 = 0.151$), and Electrode ($F(4, 120) = 7.327$, $p = 0.004$, $\eta_p^2 = 0.196$). There was no interaction between Group and Electrode ($F(4, 120) = 1.947$, $p = 0.164$, $\eta_p^2 = 0.021$). Generally, N400 elicited by the highly surprised group (Mean = 6.061; SE = 1.086) was larger than the less surprised group (Mean = 9.615; SE = 1.086). Post-hoc comparisons showed that the central site (Cz) elicited stronger negativity than central-parietal sites (CPz and Pz).

Three additional ANOVAs were also performed in the time windows of 200–250 ms, 500–700 ms, and 700–1000 ms. In the time window of 200–250 ms, the results revealed main effects of Group ($F(1, 30) = 4.581$, $p = 0.041$, $\eta_p^2 = 0.132$) and Electrode ($F(4, 120) = 23.354$, $p < 0.001$, $\eta_p^2 = 0.438$), but no interaction between the two factors ($F(4, 120) = 0.567$, $p = 0.531$,

Table 3
The mean ratios of participants' responses meeting the expected funniness ratings.

Ratio	Joke (rated as 3 or 4)		Non-joke (rated as 1 or 2)	
	Mean	SD	Mean	SD
Highly surprised group	63.66%	26.38%	94.06%	6.40%
Less surprised group	59.38%	20.33%	93.98%	10.37%
Good comprehension group	72.76%	16.31%	92.91%	10.31%
Poor comprehension group	50.27%	24.15%	95.13%	6.29%
Highly amused group	78.36%	13.62%	90.58%	10.43%
Less amused group	44.67%	18.11%	97.45%	3.81%
Total	61.52%	23.27%	94.02%	8.48%

$\eta_p^2 = 0.019$). Generally, the less surprised group (Mean = 8.878, SE = 0.901) elicited a more positive amplitude than the highly surprised group (Mean = 6.152, SE = 0.901). Post-hoc comparisons showed that ERP amplitudes at anterior sites (Fz, FCz, and Cz) were more positive than posterior sites (CPz and Pz). However, the analysis in the 500–700 ms and 700–1000 ms time window showed only a main effect of Electrode in the former time window ($F(4, 120) = 12.491$, $p < 0.001$, $\eta_p^2 = 0.294$; $F(4, 120) = 1.582$, $p = 0.211$, $\eta_p^2 = 0.050$), but not of Group ($F(1, 30) = 1.772$, $p = 0.193$, $\eta_p^2 = 0.056$; $F(1, 30) = 0.280$, $p = 0.601$, $\eta_p^2 = 0.009$), or any interaction between the two factors ($F(4, 120) = 1.218$, $p = 0.302$, $\eta_p^2 = 0.039$; $F(4, 120) = 0.761$, $p = 0.486$, $\eta_p^2 = 0.025$) (Fig. 4).

3.2.2.2. Good vs. poor comprehension group. To further verify whether the P600 component would be influenced by participants' comprehensibility to the joke stimuli, a two-way mixed-design ANOVA involving the between-subject factor of Group (good/poor comprehension group) and the within-subject factor of Electrode (Fz, FCz, Cz, CPz, Pz) was performed in the time window of 500 and 700 ms. The statistical results revealed main effects of Group ($F(1, 30) = 4.743$, $p = 0.037$, $\eta_p^2 = 0.137$) and Electrode ($F(4, 120) = 12.321$, $p < 0.001$, $\eta_p^2 = 0.291$). There was no interaction between the two factors ($F(4, 120) = 0.792$, $p = 0.451$, $\eta_p^2 = 0.026$). Generally, P600 elicited by the good comprehension group (Mean = 10.736; SE = 0.959) was larger than the poor comprehension group (Mean = 7.781; SE = 0.959). Follow-up comparisons showed that posterior sites (CPz and Pz) elicited more positive amplitudes than anterior sites (Fz, FCz, and Cz).

Three additional ANOVAs were sequentially performed in the time windows of 200–250 ms, 350–500 ms, and 700–1000 ms. The results revealed main effects of Electrode in the time windows of 200–250 ms and 350–500 ms ($F(4, 120) = 24.763$, $p < 0.001$, $\eta_p^2 = 0.452$; $F(4, 120) = 7.061$, $p = 0.005$, $\eta_p^2 = 0.191$), but not in the 700–1000 ms time window ($F(4, 120) = 1.577$, $p = 0.212$, $\eta_p^2 = 0.050$). The Group effect was only significant in the 700–1000 ms time window ($F(1, 30) = 6.100$, $p = 0.019$, $\eta_p^2 = 0.169$). Generally, the good comprehension group (Mean = 9.956, SE = 0.897) elicited a more positive amplitude than the poor comprehension group (Mean = 6.822, SE = 0.897) in the 700–1000 ms time window. No other Group effects (200–250 ms: $F(1, 30) = 2.052$, $p = 0.162$, $\eta_p^2 = 0.064$; 350–500 ms: $F(1, 30) = 1.958$, $p = 0.172$, $\eta_p^2 = 0.061$) or interactions between the two factors ($F(4, 120) = 2.412$, $p = 0.109$, $\eta_p^2 = 0.074$; $F(4, 120) = 0.787$, $p = 0.424$, $\eta_p^2 = 0.026$; $F(4, 120) = 0.677$, $p = 0.527$, $\eta_p^2 = 0.022$) were found (Fig. 5).

3.2.2.3. Highly vs. less amused group. Finally, to inspect whether participants' feeling of funniness to the joke stimuli would impact on LPP amplitudes, a two-way mixed-design ANOVA involving the between-subject factor of Group (highly/less amused group) and the within-subject factor of Electrode (Fz, FCz, Cz, CPz, Pz) was conducted in the time window of 700 and 1000 ms. The statistical results revealed a main effect of Group ($F(1, 30) = 4.459$, $p = 0.043$, $\eta_p^2 = 0.129$), but not of Electrode ($F(4, 120) = 1.590$, $p = 0.209$, $\eta_p^2 = 0.050$). There was no interaction between the two factors ($F(4, 120) = 0.928$, $p = 0.409$, $\eta_p^2 = 0.030$). Generally, the highly amused group (Mean = 9.760; SE = 0.918) elicited larger LPP amplitudes than the less amused group (Mean = 7.018; SE = 0.918).

Three additional ANOVAs were also performed in the time windows of 200–250ms, 350–500 ms, and 500–700 ms. In all the three time windows, analyses showed main effects of Electrode ($F(4, 120) = 23.131$, $p < 0.001$, $\eta_p^2 = 0.435$; $F(4, 120) = 6.975$, $p = 0.006$, $\eta_p^2 = 0.189$; $F(4, 120) = 12.319$, $p < 0.001$, $\eta_p^2 = 0.291$). No other Group effects ($F(1, 30) = 1.534$, $p = 0.225$, $\eta_p^2 = 0.049$; $F(1, 30) = 1.157$, $p = 0.291$, $\eta_p^2 = 0.037$; $F(1, 30) = 2.494$, $p = 0.125$, $\eta_p^2 = 0.077$) or interaction between the two factors ($F(4, 120) = 0.276$, $p = 0.709$, $\eta_p^2 = 0.009$; $F(4, 120) = 0.411$, $p = 0.595$, $\eta_p^2 = 0.014$; $F(4, 120) = 0.788$, $p = 0.451$, $\eta_p^2 = 0.062$) were found (Fig. 6).

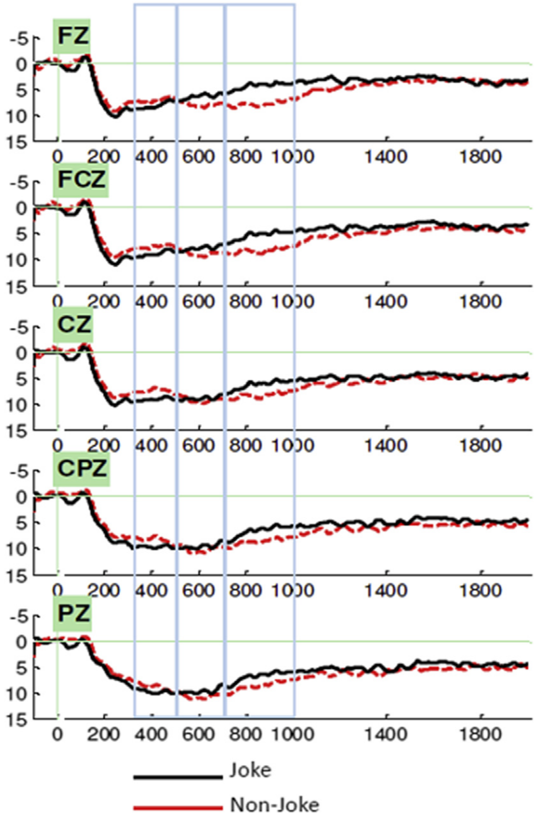

Fig. 3. Representative electrodes of ERP amplitudes elicted by jokes and non-jokes. Voltage scales in μV. Time scales in milliseconds. In comparison with non-jokes, jokes elicited stronger N400 and LPP effects in the time windows of 350–500ms and 700–1000ms, respectively. There was no differences of ERP amplitudes in the 500–700 ms time window.

4. Discussion

To elucidate the neural basis of humor processing from the perspective of temporal resolution, we utilized the ERP technique to probe the temporal dynamics of sub-stages in processing of Chinese jokes (question/answer type statements). More importantly, to investigate if the variances of readers' behavioral patterns would influence the occurrence of corresponding neurophysiological activities, we categorized the participants into highly vs. less surprised, comprehended, and amused groups according to their rating scores on jokes. Our omnibus analysis showed that jokes tended to elicit larger negativity than non-jokes did with a focus on anterior sites in the time window of 350–500 ms. The result may suggest the N400 component, an endogenous event-related brain potential which reflects increased difficulty of integrating current lexical information into previously established discourse or pragmatic contexts (Leuthold, Filik, Murphy, & Mackenzie, 2012; Nieuwland, Ditman, & Kuperberg, 2010; Van den Brink et al., 2012), can be sensitive to the detection of semantic incongruity during the initial humor processing (Du et al., 2013; Feng et al., 2014; Mayerhofer & Schacht, 2015).

In the later time window of 500–700 ms, no P600 effect was elicited by comparing jokes and non-jokes. P600 is broadly associated with reprocessing costs–reviewing previous contexts and repairing problems if processing difficulties occur (Van Petten & Luka, 2012). Such a concept could occur when participants try to get any attainable cues in the setup sentences to form possible resolution to the punch lines (Coulson & Kutas, 2001; Du et al., 2013; Feng et al., 2014; Marinkovic et al., 2011). Hence, the results consistently suggested the same degrees of resolution for jokes and non-jokes, as reflected by participants' similar comprehension scores to both types of stimuli. After approximately 700 ms post-stimulus, a broad positive potential elicited by jokes peaked at all the midline electrodes, with larger effects at anterior sites. We speculated that this may reflect LPP, a sustained positivity that may reflect further elaboration and mirth experiences, as it can be modulated by emotional relevance (Liu, Huang, McGinnis-Deweese, Keil, & Ding, 2012; Sabatinelli, Keil, Frank, & Lang, 2013) and positively correlated with emotional arousal (Gierych, Milner, & Michalski, 2005). These results re-verified the cognitive functions involved in humor processing with different electrophysiological activities. The results of the between-group analysis were discussed in the following subsections.

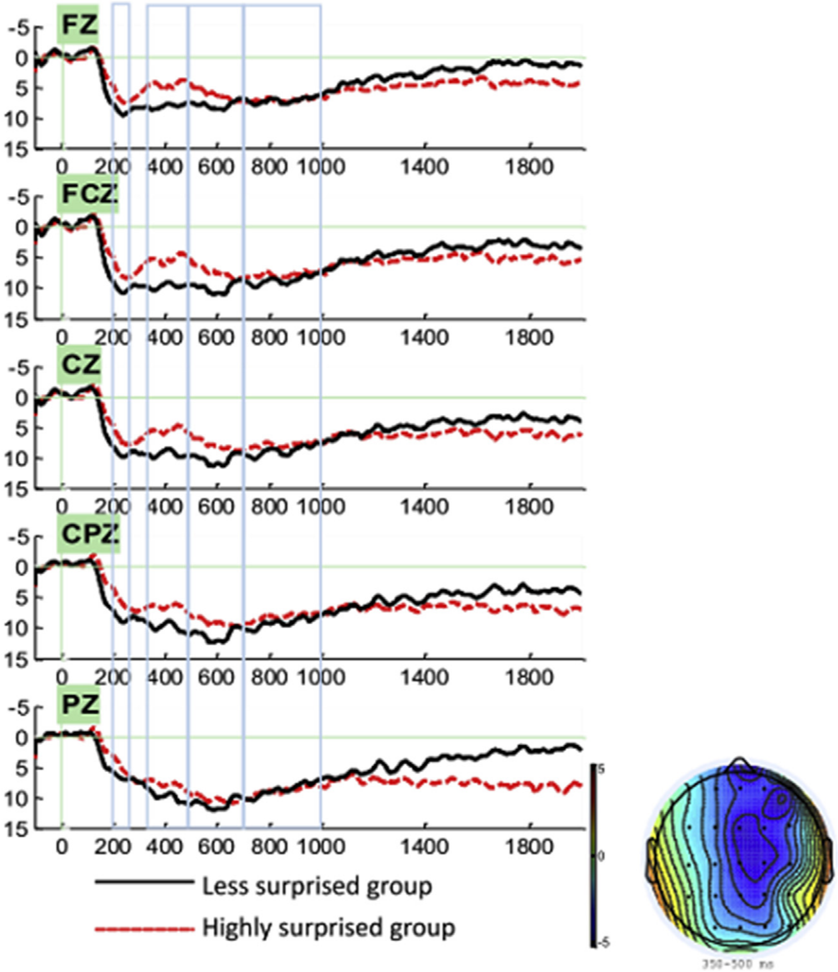

Fig. 4. The comparison waves of jokes elicited by the highly and less surprised group, and the topograpghy of the difference wave (highly surprised group − less surprised group) in the time window of 350–500 ms. Voltage scales in μV. Time scales in milliseconds.

4.1. Surprise

Based on the three-stage model, we suggested that participants' surprise feelings could reflect the cognitive process of incongruity detection in humor processing. In line with our hypothesis, the highly surprised group elicited a larger N400 in reading jokes compared with the less surprised one, and the N400 effect was stronger at more anterior sites. These results probably revealed a larger mismatch or incongruity between our setup sentences and punch lines perceived in the highly surprised group. According to the previous research (Moreno & Rivera, 2014; Van Petten & Luka, 2012), N400 amplitudes may reflect the match of the global context and the target words presented in terms of semantic expectations, and unexpected outcomes would lead to larger N400s than highly expected ones. It is the incongruent expectation shaped by the setup sentences of jokes that can contribute to surprise during individuals' perception of humor (Suls, 1972). The larger violation of the expectation built from the setup sentences may thus be registered in the highly surprised group, and reflected by larger N400 amplitudes.

The highly surprised group also showed a smaller P2 compared with the less surprised group. As noted in the previous study (Federmeier & Kutas, 2002; Hillyard & Muente, 1984; Luck & Hillyard, 1994), the P2 effect was mostly related to visual feature processing during visual search with pictorial stimuli. In language processing, P2 could be modulated by the contextual information such as sentence-level constraints or congruity associated with the target word (Coulson & Brang, 2010; Coulson, Federmeier, Van Petten, & Kutas, 2005; Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007). Larger P2 amplitudes were observed for expected endings than for unexpected ones in the sentence reading task (Moreno & Rivera, 2014). Hence, the larger P2 effect in the less surprised group might reflect an early influence of high expectancy built up by concise punch lines in the jokes. Alternatively, it seemed plausible that the larger P2 could be driven by the subsequent

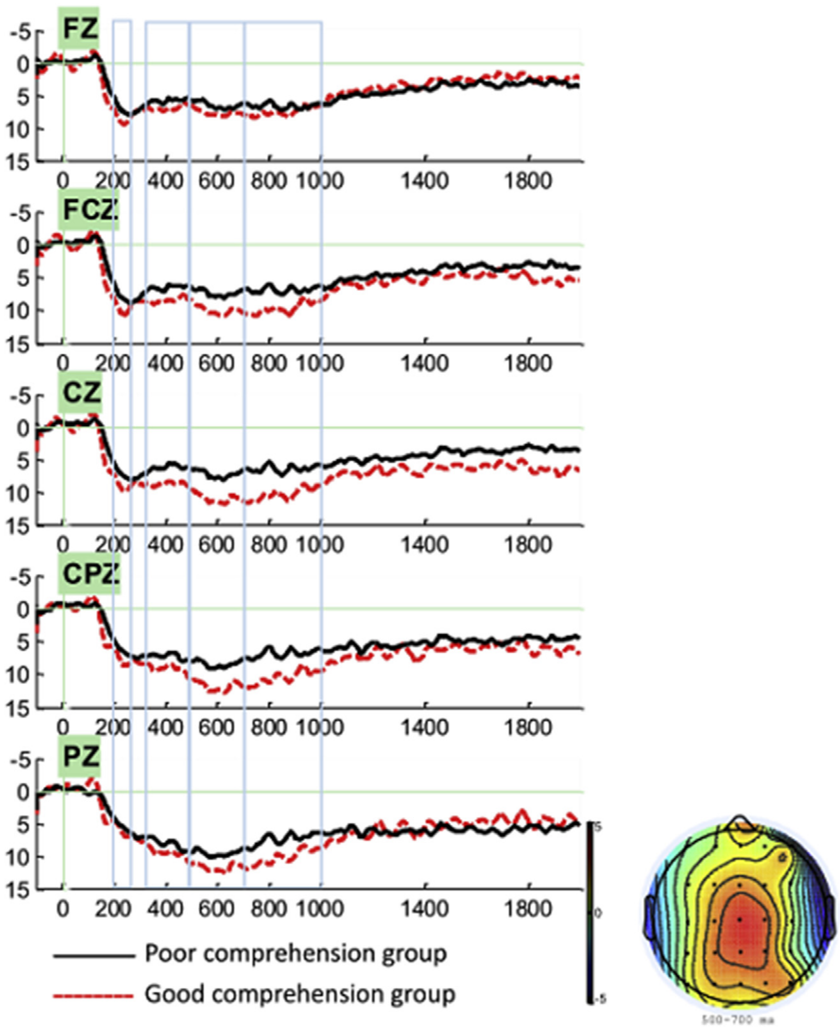

Fig. 5. The comparison waves of jokes elicited by the good and poor comprehension group, and the topograpghy of the difference wave (good comprehension group − poor comprehension group) in the time window of 500–700 ms. Voltage scales in μV. Time scales in milliseconds.

smaller N400 of opposite polarity for the less surprised group. We thus suggested that N400 would still be a primary biomarker for incongruity detection of humor processing based on the N400 effects in both the omnibus and between-group analysis.

4.2. Comprehension

According to the three-stage model, we suggested that it is the comprehensibility that would reflect the cognitive process of incongruity resolution in humor processing. To our expectation, the good comprehension group showed a greater posterior P600 in reading jokes. Coulson and Kutas (2001) has found the posterior P600 in good comprehenders when reading high-constraint jokes, possibly suggesting the violation of expectation built from setup sentences. In line with recent studies (Du et al., 2013; Feng et al., 2014; Marinkovic et al., 2011), we postulated that it could further reflect good comprehenders' successful attempt in re-interpreting novel associations to resolve the incongruity at punch lines. On the contrary, poor comprehenders failed to obtain attainable interpretations in forming possible resolution to the punch lines. More attention resource may thus be needed in poor comprehenders, leading to their weaker P600 amplitudes (Kok, 1997; Polich, 2007).

Aside from the P600 effect, the good comprehension group also displayed a larger sustained positivity than the poor comprehension group did between 700 and 1000 ms post-stimuli at the midline electrodes. Based on Mayerhofer and Schacht's (2015) study, the sustained positivity in this time window could suggest participants' extensive elaboration of the humorous stimuli. As supported by our behavioral data, the good comprehension group generally scored higher on the funniness ratings of jokes. These results further emphasized a more important role of comprehensibility rather than surprise

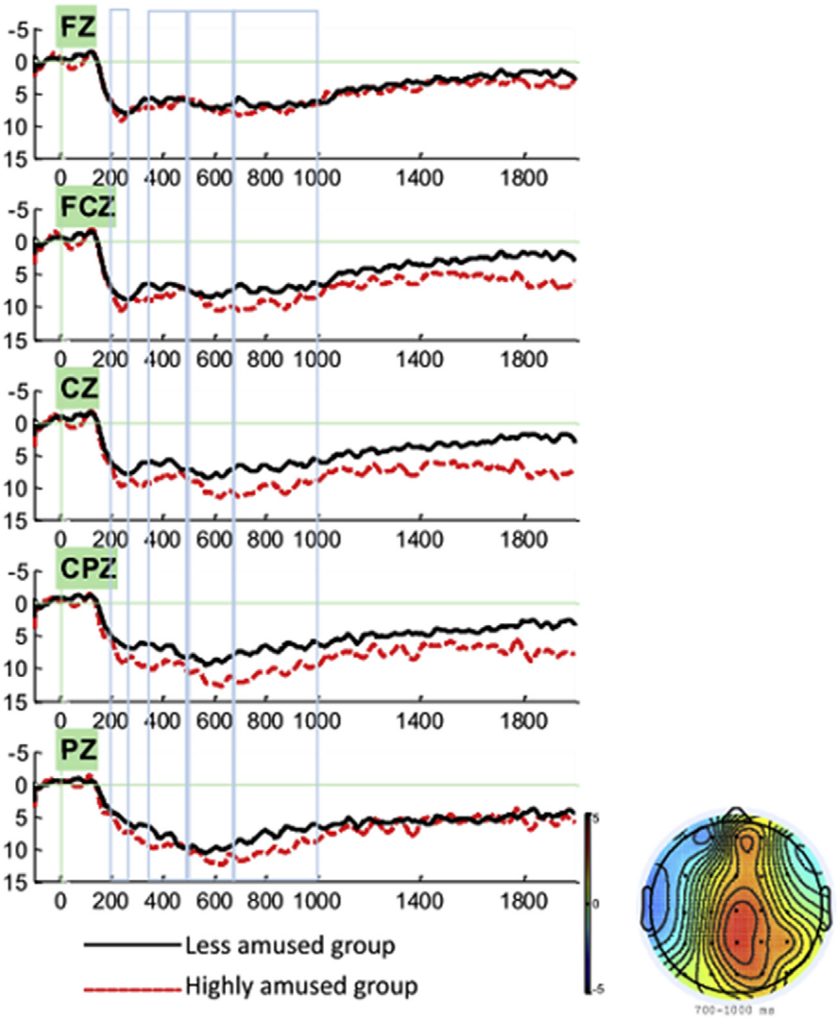

Fig. 6. The comparison waves of jokes elicited by the highly and less amused group, and the topograpghy of the difference wave (highly amused group − less amused group) in the time window of 700–1000 ms. Voltage scales in μV. Time scales in milliseconds.

in "getting" humor with our question/answer type humorous stimuli. Therefore, we suggested that P600 would still be a valid indicator in identifying the stage of incongruity resolution based on the observation of P600 effects elicited by people with different comprehension levels of jokes.

4.3. Funniness

Following the three-stage model, we suggested that a feeling of funniness (amusement) could be invoked by further inference in the cognitive process of elaboration during humor processing (Wyer & Collins, 1992). To our expectation, the highly amused group elicited stronger LPPs at the midline electrodes than the less amused group in reading jokes. This effect may imply the highly amused group's more affective elaboration of the jokes, as supported by their higher ratios in categorizing jokes as funny stimuli. Moreover, with emotionally biased and unbiased sentences, Moreno and Rivera's (2014) study has argued that the post-N400 late positivity and its topographic distribution (frontal or posterior) could be associated with degrees of conflict between unexpected and expected linguistic events and/or emotional connotations of unexpected yet plausible endings. We therefore postulated that our results could indicate the highly amused group's controlled evaluative processes in interpreting and sustained attention towards the unexpected yet funny riddle punch lines, due to the more widespread topographical distribution of the effect compared with the posterior P600. Altogether, these results showed that LPP would still be an effective index for identifying the stage of elaboration from the observation of groups with different levels of amusement in humor processing.

In summary, for the between-group differences in surprise ratings to the jokes, which may be related to the initial incongruity-detection stage of humor processing, ERP waves showed a main N400 effect (350–500 ms), with an early deflection in the time window of 200–250 ms. In addition, participants with different ratings of comprehension for jokes exerted a primary influence on ERPs in the later time window of P600 (500–700 ms), with a sustained deflection between 700 and 1000 ms. Lastly, people with different levels of amusement evoked by jokes revealed a sustained LPP effect (700–1000 ms) as predicted. Consequently, our ERP findings not only affirmed how the three sub-stages of humor processing were differentiated from each other, but it also showed the stability of the three-stage model by examining participants' surprise, comprehensibility, and amusement levels in reading jokes with the three corresponding electrophysiological indices of N400, P600, and LPP.

Although the current study validated the three-stage model of humor processing according to Feng et al. (2014), several limitations and further refinements should be considered before reaching solid explanation towards the impact of intergroup differences on humor processing. First, the use of EEG allows a detailed inspection of temporal dynamics involved in humor processing. However, the lack of better source localization may render different possibilities of the observed late effects and this could be resolved by supplement data with magnetoencephalography (MEG), which possesses higher spatial resolutions. In addition, although we aimed to investigate readers' variance in processing humor, other variances in materials such as the degree of plausibility and participants' familiarity towards jokes could also influence the behavioral patterns and neural activities observed in the present study. Therefore, a better evaluation and careful control of experimental materials in future research may offer more insight into the unseen effects caused by between-subject differences in humor processing.

Acknowledgements

This research was partially supported by the “Aim for the Top University Project” of National Taiwan Normal University (NTNU), sponsored by the Ministry of Education, Taiwan, R.O.C. and the “International Research-Intensive Center of Excellence Program” of NTNU and Ministry of Science and Technology, Taiwan, R.O.C. under Grant no. MOST 104-2911-I-003-301.

References

Alden, D. L., Hoyer, W. D., & Lee, C. (1993). Identifying global and culture-specific dimensions of humor in advertising: A multinational analysis. *The Journal of Marketing*, 64–75.
Chan, Y.-C., Chou, T.-L., Chen, H.-C., & Liang, K.-C. (2012). Segregating the comprehension and elaboration processing of verbal jokes: An fMRI study. *NeuroImage, 61*(4), 899–906.
Chan, Y.-C., Chou, T.-L., Chen, H.-C., Yeh, Y.-C., Lavallee, J. P., Liang, K.-C., et al. (2013). Towards a neural circuit model of verbal humor processing: An fMRI study of the neural substrates of incongruity detection and resolution. *NeuroImage, 66*, 169–176.
Coulson, S., & Brang, D. (2010). Sentence context affects the brain response to masked words. *Brain and language, 113*(3), 149–155.
Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and cognitive processes, 13*(1), 21–58.
Coulson, S., Federmeier, K. D., Van Petten, C., & Kutas, M. (2005). Right hemisphere sensitivity to word-and sentence-level context: Evidence from event-related brain potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(1), 129.
Coulson, S., & Kutas, M. (2001). Getting it: Human event-related brain response to jokes in good and poor comprehenders. *Neuroscience letters, 316*(2), 71–74.
Donchin, E. (1981). Surprise! … surprise? *Psychophysiology, 18*(5), 493–513.
Du, X., Qin, Y., Tu, S., Yin, H., Wang, T., Yu, C., et al. (2013). Differentiation of stages in joke comprehension: Evidence from an ERP study. *International Journal of Psychology, 48*(2), 149–157.
Federmeier, K. D., & Kutas, M. (2002). Picture the difference: Electrophysiological investigations of picture processing in the two cerebral hemispheres. *Neuropsychologia, 40*(7), 730–747.
Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain research, 1146*, 75–84.
Feng, Y.-J., Chan, Y.-C., & Chen, H.-C. (2014). Specialization of neural mechanisms underlying the three-stage model in humor processing: An ERP study. *Journal of Neurolinguistics, 32*, 59–70.
Gierych, E., Milner, R., & Michalski, A. (2005). ERP responses to smile-provoking pictures. *Journal of Psychophysiology, 19*(2), 77–90.
Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika, 24*(2), 95–112.
Hamm, J. P., Johnson, B. W., & Kirk, I. J. (2002). Comparison of the N300 and N400 ERPs to picture stimuli in congruent and incongruent contexts. *Clinical Neurophysiology, 113*(8), 1339–1350.
Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis: Early automatic and late controlled processes. *Journal of cognitive neuroscience, 11*(2), 194–205.
Hehl, F.-J., & Ruch, W. (1985). The location of sense of humor within comprehensive personality spaces: An exploratory study. *Personality and Individual Differences, 6*(6), 703–715.
Hillyard, S. A., & Münte, T. F. (1984). Selective attention to color and location: An analysis with event-related brain potentials. *Perception & psychophysics, 36*(2), 185–198.
Kohn, N., Kellermann, T., Gur, R. C., Schneider, F., & Habel, U. (2011). Gender differences in the neural correlates of humor processing: Implications for different processing modes. *Neuropsychologia, 49*(5), 888–897.
Kok, A. (1997). Event-related-potential (ERP) reflections of mental resources: A review and synthesis. *Biological Psychology, 45*(1), 19–56.
Leuthold, H., Filik, R., Murphy, K., & Mackenzie, I. G. (2012). The on-line processing of socio-emotional information in prototypical scenarios: Inferences from brain potentials. *Social cognitive and affective neuroscience, 7*(4), 457–466.
Liu, Y., Huang, H., McGinnis-Deweese, M., Keil, A., & Ding, M. (2012). Neural substrate of the late positive potential in emotional processing. *The Journal of Neuroscience, 32*(42), 14563–14572.
Luck, S. J., & Hillyard, S. A. (1994). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology, 31*(3), 291–308.
Marinkovic, K., Baldwin, S., Courtney, M. G., Witzel, T., Dale, A. M., & Halgren, E. (2011). Right hemisphere has the last laugh: Neural dynamics of joke appreciation. *Cognitive, Affective & Behavioral Neuroscience, 11*(1), 113–130.
Mayerhofer, B., & Schacht, A. (2015). From incoherence to mirth: Neuro-cognitive processing of garden-path jokes. *Frontiers in psychology, 6*.
Mobbs, D., Hagan, C. C., Azim, E., Menon, V., & Reiss, A. L. (2005). Personality predicts activity in reward and emotional regions associated with humor. *Proceedings of the National Academy of Sciences of the United States of America, 102*(45), 16502–16506.
Moratti, S., Saugar, C., & Strange, B. A. (2011). Prefrontal-occipitoparietal coupling underlies late latency human neuronal responses to emotion. *The Journal of Neuroscience, 31*(47), 17278–17286.

Moreno, E. M., & Rivera, I. C. (2014). Setbacks, pleasant surprises and the simply unexpected: Brainwave responses in a language comprehension task. *Social cognitive and affective neuroscience, 9*(7), 991–999.

Nieuwland, M. S., Ditman, T., & Kuperberg, G. R. (2010). On the incrementality of pragmatic processing: An ERP investigation of informativeness and pragmatic abilities. *Journal of Memory and Language, 63*(3), 324–346.

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology, 118*(10), 2128–2148.

Sabatinelli, D., Keil, A., Frank, D. W., & Lang, P. J. (2013). Emotional perception: Correspondence of early and late event-related potentials with cortical and subcortical functional MRI. *Biological psychology, 92*(3), 513–519.

Suls, J. M. (1972). A two-stage model for the appreciation of jokes and cartoons: An information-processing analysis. *The psychology of humor: Theoretical perspectives and empirical issues, 1*, 81–100.

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology, 83*(2), 176–190.

Van den Brink, D., Van Berkum, J. J., Bastiaansen, M. C., Tesink, C. M., Kos, M., et al. (2012). Empathy matters: ERP evidence for inter-individual differences in social language processing. *Social cognitive and affective neuroscience, 7*(2), 173–183.

Wyer, R. S., & Collins, J. E. (1992). A theory of humor elicitation. *Psychological review, 99*(4), 663.