

Talker's voice and gender stereotype in human auditory sentence processing – evidence from event-related brain potentials

Sonja Lattner*, Angela D. Friederici

Max-Planck-Institute of Cognitive Neuroscience, Stephanstrasse 1a, 04103 Leipzig, Germany

Received 7 November 2002; received in revised form 20 December 2002; accepted 20 December 2002

Abstract

The present study investigated the influence of implicit speaker information on the sentence interpretation. We auditorily presented sentences that comprised of either stereotypically male or stereotypically female self-referent utterances. In the congruent conditions, these utterances were produced by speakers whose gender matched the semantic content. In the incongruent condition, stereotypically male utterances were produced by female speakers and vice versa. The event-related brain potentials (ERP) of 32 listeners exhibited a late positivity (P600) for the incongruent condition. No significant differences were observed between male and female listeners. In the absence of any ERP effect in the earlier time range, it was concluded that the access of the semantic information as such is independent of the speaker's voice, but that speaker property, semantic content and stereotypical knowledge are integrated in a later processing stage. © 2003 Elsevier Science Ireland Ltd. All rights reserved.

Keywords: Event-related brain potentials; Speaker information; Stereotypes; Speaker gender; Late positivity

In speech communication the listener, not only decodes the speaker's intended linguistic message from the acoustic signal, but at the same time he/she extracts information about age, gender and other properties of the speaker. In the discourse situation, this implicit information serves as part of the context knowledge. The aim of the present experiment was to investigate how this information influences the interpretation of an utterance.

Previous studies registering event-related brain potentials (ERPs) while subjects were presented with sentences, have shown that a violation of explicit context knowledge about a person leads to a distinct brain response. In one study [3], subjects had learned facts about fictitious people (e.g. Mary is a lawyer). In the test session they were presented with facts confirming or contradicting these learned facts. A contradicting target utterance (e.g. Mary is a chemist) led to a negative ERP called N400 which is event-locked to the presentation of the contradicting word. Similar effects have been reported in relation to the preceding discourse information (e.g. 'Jane told her brother that he was exceptionally slow', presented in a context where he was described as being very fast) [14]. As the

N400 is typically interpreted as reflecting difficulties in semantic/pragmatic integration [1,6,9–11,15] it was concluded from the latter result that rapid word integration is influenced by a broad range of context factors including explicit knowledge as well as the discourse information.

However, another study demonstrated that a target word violating stereotypical assumptions (e.g. The driver of the wrecked car pulled herself through the window) does not lead to an N400 effect but to a late positive deflection (P600) of the ERP [12]. The P600 is to date assumed to reflect processes of repair or reanalysis, particularly during grammatical violations [4,5,7,8]. It has therefore been suggested that the P600 effect observed in response to stereotype violations may reflect similar processes involving re-integration of semantic meaning and stereotypical beliefs [12]. Note, however, that the P600 effect has only been demonstrated for reflexive pronouns, i.e. function words, whereas the target words in the other studies consisted of content words.

Unlike the earlier studies, the present experiment used acoustic stimulation. This allowed the role of implicit speaker information in semantic/pragmatic integration of words to be investigated. The questions in the focus of the present study were: (1) Does speaker information establish a semantic/pragmatic context, and does a violation of the

* Corresponding author. Fax: +49-341-99-40-279.

E-mail address: lattner@cns.mpg.de (S. Lattner).

stereotypical information established by this context lead to a distinct event-related potential?; (2) If so, does the incongruity of the speaker context and stereotypical beliefs lead to a distinct response at the earlier time range of the ERP (e.g. an N400) or later time range (e.g. a P600)?

The stimuli consisted of 280 spoken utterances that were divided into two conditions labelled *congruent* and *incongruent*. The congruent condition consisted of 140 sentences. Seventy of these sentences described a self-referent property that is stereotypical for males (e.g. 'I like to play soccer'). These sentences were uttered by four male speakers, whereby each speaker read 17–18 sentences. The remaining 70 sentences consisted of self-referent statements that described a stereotypically female property (e.g. 'I like to wear lipstick'). These sentences were uttered by four female speakers, whereby each speaker read 17–18 sentences. The incongruent conditions comprised of the same 140 sentences as the congruent condition, but in this case, the self-referent statements about stereotypically male properties were uttered by a female speaker, and vice versa. The intensity of the stimuli was normalized across conditions using the normalizing tool of the Cool Edit speech editing software. Note that the congruity or incongruity of the sentences only depended on the sentence final word.

The stimuli were rated by 24 undergraduate students (12 male), who did not participate in the electroencephalograph (EEG) experiment. Each student rated 140 sentences in order to avoid repetitions of items. The rating was performed on a five point scale (−2: very untypical; +2 very typical), with one block containing sentences uttered by male speakers, and the other block containing sentences rated by female speakers. The block order was counter-balanced across raters. The congruent sentences gained an average of 1.33 points (typically male: 1.33; typically female: 1.33); the incongruent sentences gained an average rating of −0.77 points (untypically male: −0.76 untypically female: −0.79). The statistical analysis (ANOVA of the factors speaker gender, rater gender, congruity) revealed no significant differences in rating values dependent on the gender of the speaker, nor between the male or female raters. The only significant difference was due to the factor congruity ($F_{3,45} = 4.77, P < 0.01$), confirming the validity of the stimulus material.

Thirty-two healthy, normal hearing, right handed native speakers of German served as subjects (16 male, age 21–30 yrs; 16 female, age 20–30 yrs.; both groups were matched in terms of handedness and Reading Span [2]) served as subjects in the ERP experiment. They were undergraduate students and were paid for participation and gave written informed consent in accordance with the guidelines approved by the Ethics Committee of the Leipzig University Medical Faculty and the Declaration of Helsinki. The subjects were tested individually. They were presented with the stimuli in a sound proof EEG cabine using the ERTS presentation software (Experimental Runtime System,

Berisoft Corporation 1995). A trial started with a 1200 ms prestimulus presentation of a fixation cross. While the fixation cross remained on the screen the stimulus was presented via headphones (Hanuma Tempest Digital ProH700) at a comfortable volume of about 70 dB SPL. After the stimulus offset, the fixation cross remained for another 1200 ms. After a 400 ms blank screen, subjects were instructed to blink during a pause of 3500 ms, before the next trial began. A bipolar electrooculogram was used to reject artefacts from blinking during sentence presentation. In order to maintain subjects' attention, 16 trials (two of each speaker) were followed by a content question. The subjects' brain responses were recorded at a sampling frequency of 250 Hz using AG + /AGCL- electrodes and a TMS International Amplifier. The left mastoid (A1) served as reference electrode during the acquisition; offline, the ERPs were re-referenced against the linked mastoids and filtered by a 45 Hz lp-filter. The ERPs in response to the sentence final word were compared for the congruent conditions vs. incongruent conditions. The ERP onset was at the beginning of the sentence-final word with a −200 to 0 ms pre-target baseline. As the sentences were identical in each condition, the frequency of occurrence as determined by the CELEX database and cloze probability of the sentence final words were balanced across conditions.

As Fig. 1 shows, the average ERPs of congruent and incongruent conditions differ by a positivity at posterior electrode sites for the incongruent condition. The effect started around 600 ms and was largest at the PO8 electrode, where it exhibited a peak latency of 832 ms and an amplitude of 1.3 μ V. It is henceforth labelled P600 effect. A statistical analysis was performed over the posterior electrodes (TP7, P7, P3, PO7, CP3, PZ, OZ, CPZ, TP8, P8, P4, PO8, CP4) in the time range of 600–1000 ms. An analysis of variance (ANOVA) was performed with the factor 'congruity' (congruent vs. incongruent) and the gender of the listener as a between subject factor. The ANOVA confirms the significance of the observed P600 effect of congruity ($F_{1,30} = 9.91, P < 0.01$). In addition, there was a significant effect of listener gender ($F_{1,30} = 4.24, P < 0.05$), indicating a generally stronger ERP response for female subjects. However, there was no significant interaction ($F < 1$) of the factors listener gender and congruity, indicating that the P600 effect was equally strong in male and female subjects. Further analyses revealed no other significant effect in the ERPs.

In the present study, we asked whether context information conveyed by the speaker's voice leads to the establishment of gender stereotypical expectations and whether violations of these expectations would lead to a distinct ERP. We presented self-referent utterances that were either congruent or incongruent with common gender stereotypes. Comparing the brain responses to the sentence final noun that violated the listeners' stereotypical expectations to the responses to matching nouns revealed a late positivity in response to the incongruent utterances. Thus,

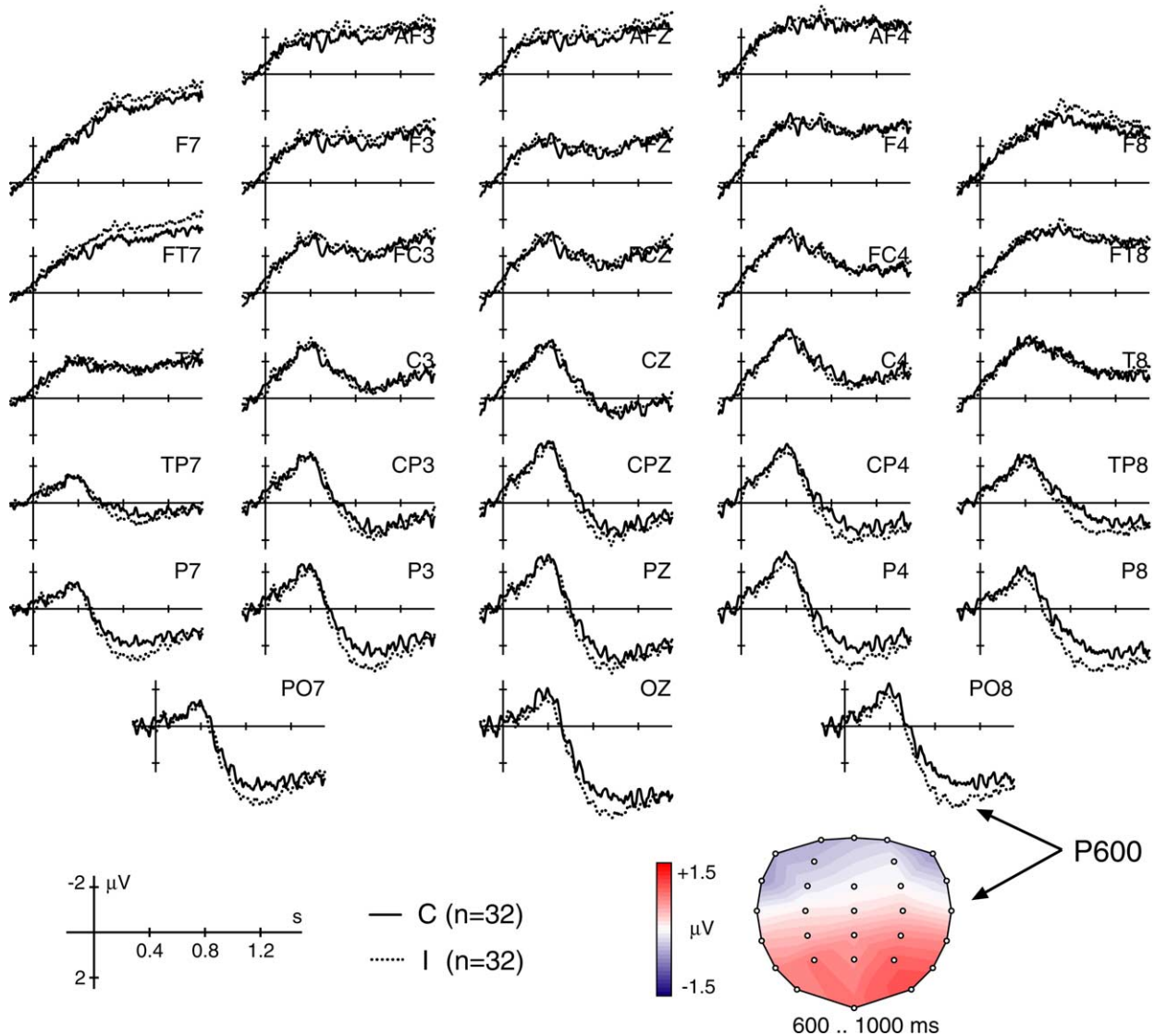


Fig. 1. Brain potentials event-locked to the presentation of the sentence final words ($n = 32$ subjects; electrode labels correspond to the guidelines of the American Electroencephalographic Society [14]; negativity plotted upwards). The solid line represents responses to congruent words (C), i.e. to words that represent a match of gender stereotypes and speaker's voice (e.g. female voice uttering 'I like to wear skirts'); the dotted line depicts responses to incongruent words (I), (e.g. male voice uttering 'I like to wear lipstick'). The ERP map (right bottom) displays the difference in the brain responses to the brain activity of incongruent – congruent conditions, averaged from 600 to 1000 ms, indicating the posterior scalp distribution of the P600 effect.

the context information conveyed by the speaker's voice seems to have an impact on the content processing similar to what has been reported for pronouns that represented either matching or mismatching stereotypes of occupational nouns [12]. In contrast to this earlier study [12], we observed no significant effect differences dependent on the gender of the listener, indicating that the strength of stereotype driven expectations was equally strong in male and female subjects. However, this null-effect may well depend on age, education, cultural background and other sample-specific factors, in the sense that gender differences for stereotypical information diminishes the younger the subjects are and the more liberal their educational and cultural background is. The most interesting result with respect to the questions in focus is the finding that the

integration of speaker voice and stereotypical knowledge takes place rather late, reflected in the P600, and is thus different from context mismatches of explicit knowledge, which lead to an N400 [3,15].

A late positivity (P600) has also been observed for syntactic processing, where it is typically assumed to reflect a difficulty in the structural reanalysis or repair [4,8,13]. However, this syntactic P600 effect is more centrally distributed than the effect observed in the present study. The difference in the distribution suggests at least partly different neural generators and requires a different functional explanation.

It has been suggested that a posterior P600 effect may reflect a 're-integration of semantic meaning and stereotypical beliefs' [12]. The present study supports this notion

and extends it to the domain of extra-linguistic, speaker information conveyed by the talker's voice.

References

- [1] J. Conolly, N. Phillips, Event-related potential components reflect phonological, semantic processing of the terminal words of spoken sentences, *J. Cogn. Neurosci.* 6 (1994) 256–266.
- [2] M. Danemann, P. Carpenter, Individual differences in working memory and reading, *J. Verbal Learn. Verbal Behav.* 19 (1980) 450–466.
- [3] I. Fischler, D. Childers, T. Acharyapaopan, N.J. Perry, Brain potentials during sentence verification – automatic aspects of comprehension, *Biol. Psychol.* 21 (1985) 83–105.
- [4] A. Friederici, The time course of syntactic activation during language processing: a model based on neuropsychological and neurophysiological data, *Brain Lang.* 50 (1995) 259–281.
- [5] A. Friederici, Towards a neural basis of auditory sentence processing, *Trends Cogn. Sci.* (2002) 78–84.
- [6] A. Friederici, E. Pfeifer, A. Hahne, Event-related potentials during natural speech processing: effects of semantic morphological and syntactic violations, *Cogn. Brain Res.* (1993) 183–192.
- [7] T. Gunter, L. Stowe, G. Mulder, When syntax meets semantics, *Psychophysiology* 34 (1997) 660–676.
- [8] P. Hagoort, C. Brown, J. Groothusen, The syntactic positive shift as an ERP measure of syntactic processing, *Lang. Cogn. Proc.* 8 (1993) 337–364.
- [9] P. Holcomb, H. Neville, Natural speech processing: an analysis using event-related brain potentials, *Psychobiology* 19 (1991) 286–300.
- [10] M. Kutas, K. Federmeier, Electrophysiology reveals semantic memory use in language comprehension, *Trends Cogn. Sci.* 4 (2000) 463–470.
- [11] M. Kutas, S. Hillyard, Event-related brain potentials to semantically inappropriate and surprisingly large words, *Biol. Psychol.* 11 (1980) 99–116.
- [12] L. Osterhout, M. Bersick, J. McLaughlin, Brain potentials reflect violations of gender stereotypes, *Mem. Cogn.* 25 (1997) 273–285.
- [13] L. Osterhout, P. Holcomb, D. Swinney, Brain potentials elicited by gardenpath sentences: evidence of the application of verb information during parsing, *J. Exp. Psychol. Learn. Mem. Cog.* 20 (1994) 786–803.
- [14] J. Van Berkum, P. Haagoort, C. Brown, Semantic integration in sentences and discourse: evidence from the N400, *J. Cogn. Neurosci.* 11 (1999) 657–671.
- [15] C. Van Petten, S. Coulson, S. Rubin, E. Plante, M. Parks, Time course of word identification and semantic integration, *J. Exp. Psychol. Learn. Mem. Cogn.* 25 (1999) 394–417.