Cross-modal integration of affective facial expression and vocal prosody

an EEG study

Weed, Ethan; Christensen, Peer

2011

Cross-modal integration of affective facial expression and vocal prosody: an EEG study

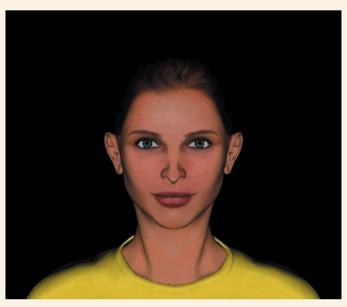
Ethan Weed[a,b] and Peer Christensen[c]

Aarhus University (a) Dept. of Linguistics, (b) Center for Functionally Integrative Neuroscience, (c) Center for Semiotics

Introduction

We have all experienced how a telephone conversation can be more challenging than speaking face to face. Understanding the intended meaning of a speaker's words requires forming an impression of the current mental state of the speaker, including her beliefs, intentions, and emotional state (Sperber & Wilson, 1995). Facial expressions are an important source of this information.

In this study, we wondered at what point emotional information from faces was integrated with the auditory processing stream. We hypothesized that the N400 component, which is sensitive to meaning at a variety of levels (Lau, Phillips, & Poeppel, 2008; Van Berkum, Van Den Brink, Tesink, Kos, & Hagoort, 2008), would be affected by incongruous emotions in face/voice pairs.



Methods

To test this, we used EEG to record brain responses to emotionally congruous and incongruous face/voice pairs. Participants viewed faces showing either a happy or a sad expression, or scrambled face. As they viewed the faces, participants heard a variety of spoken utterances delivered in either a happy or a sad tone of voice. The faces and voices were attended to passively.

Participants

Participants were 9 native speakers of Danish (4 females, 5 males, mean age 24.3). All were right-handed, and all had normal hearing and normal or corrected to normal vision (self-reported). All were undergraduates at the University of Aarhus, and received payment for their participation.

Stimuli Construction

The auditory stimuli consisted of 8 Danish sentences, recorded by a professional actor and actress. Each sentence was recorded separately by the male speaker and the female speaker and each sentence was recorded in a happy and a sad tone of voice. The visual stimuli consisted of two avatars (one male, one female) each with a happy and a sad expression. The avatars were created using Poser 6 software (e-frontier).

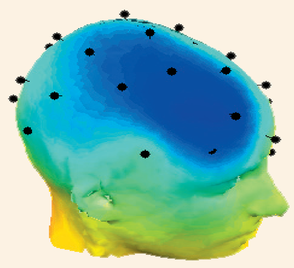
Procedure

The presentation of the stimuli was based on the paradigm used by de Gelder et. al. (1999). Following an initial fixation cross, visual stimuli were presented in the middle of a computer screen. After variable delay of between 795 and 1200 ms., a voice stimulus was played with a prosody that either matched or did not match the emotional valence of the face.

EEG acquisition and preprocessing

EEG was acquired using 32 Ag/AgCl active electrodes (actiCAP®, Brain Products, Gilching, Germany), and referenced to the tip of the nose. Resistance was kept below 5kΩ. Continuous data were filtered offline using a Butterworth bandpass filter (low cutoff: 1 Hz, high cutoff 20 Hz). Continuous data were epoched and averaged using the EEGLAB toobox (Delorme & Makeig, 2004) for MATLAB (The Mathworks, Natick, MA).

Results



N100

The N100 was measured as the mean amplitude in the difference waves at the Cz electrode within a measurement window of 100 - 150 ms. The auditory N100 component was significantly larger following an emotionally incongruent facial expression than a scrambled face (paired t-test: $t(8) = -2.92$, $p = 0.019$).

N400

The N400 was measured as the mean amplitude in the difference waves at the Cz electrode within a measurement window of 450 - 530 ms. While both the N400 waveforms in the emotionally incongruent and scambled-face conditions differed from the congruent condition, the emotionally incongruent condition did so significantly more (paired t-test: $t(8) = 3.40$, $p = 0.009$).

Scalp distribution

Both the N100 and the N400 components were most prominent at central and frontal midline electrodes.

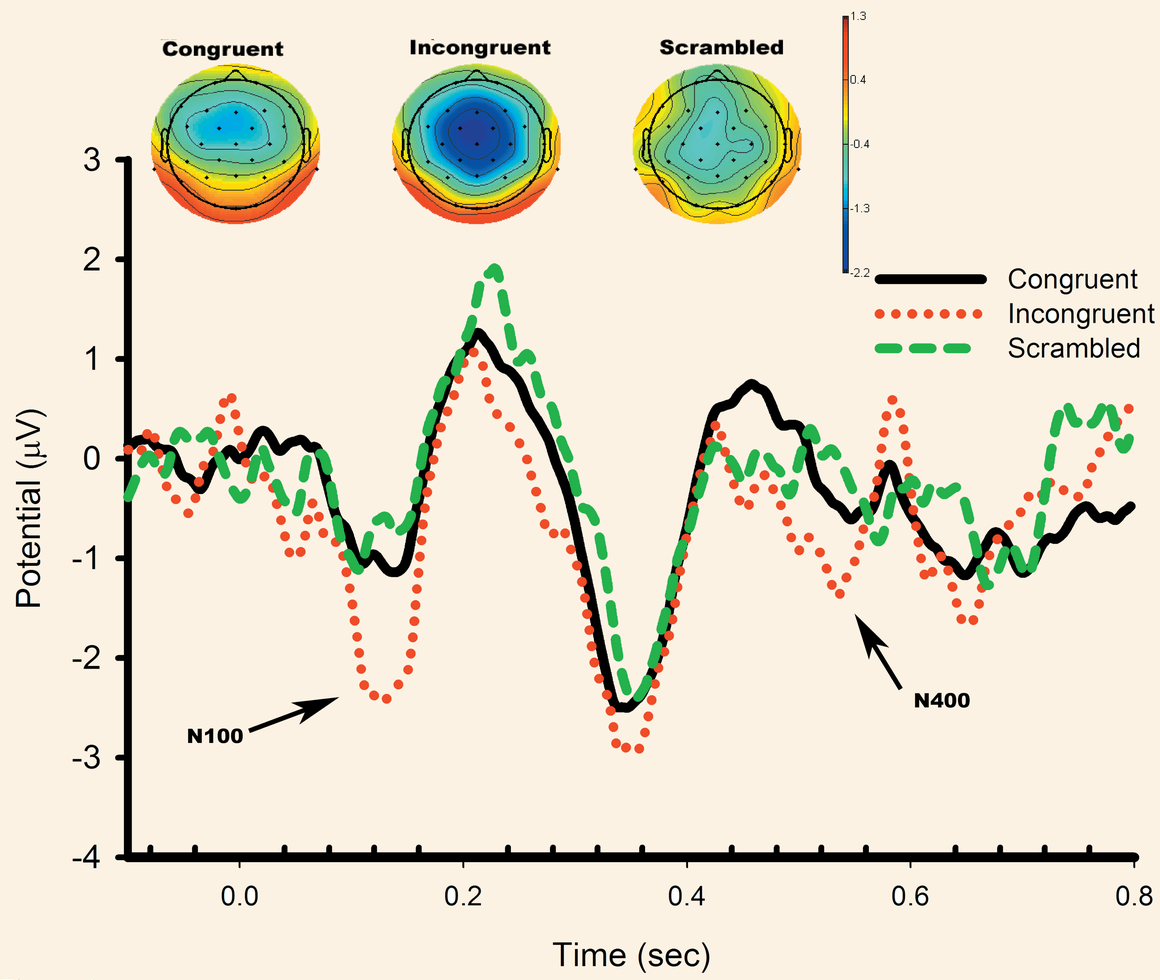


Figure 1:
Waveforms for the three conditions (emotionally congruent, emotionally incongruent, and scambled face) as recorded at the Cz electrode. Scalp maps show distribution of the N100 component.

Conclusion

Our results show that as early 100 msec after onset of spoken utterances, the brain has made an initial comparison of the affect expressed by the speaker's facial expression, and that expressed by vocal prosody. This suggests that early multi-modal perceptual areas are involved in computations which may be critical to interpretation of speaker meaning, and that integration of face/voice affective information takes place long before an utterance is completed. We suggest that the N400-like effect observed at frontal-central electrodes is a reflection of an attempted integration of incongruent information in the construction of speaker meaning.

References

de Gelder, B., Böcker, K. B. E., Tuomainen, J., Hensen, M., & Vroomen, J. (1999) The combined perception of emotion from voice and face: early interaction revealed by human electric brain responses. Neuroscience Letters 260 (1999) 133–136.

Lau, E., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics:(de) constructing the N400. *Nature Reviews Neuroscience, 9*(12), 920-933.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition*: Blackwell Pub.

Van Berkum, J., Van Den Brink, D., Tesink, C., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience, 20*(4), 580-591.