Uhm… What's going on? An EEG study on perception of filled pauses in spontaneous Swedish speech

Sebastian Mårback[1], Gustav Sjöberg[1], Iris-Corinna Schwarz[1] and Robert Eklund[2, 3]
[1]Dept of Linguistics, Stockholm University, Stockholm, Sweden
[2]Dept of Clinical Neuroscience, Karolinska Institute/Stockholm Brain Institute, Stockholm, Sweden
[3]Voice Provider Sweden, Stockholm, Sweden

Abstract

Filled pauses have been shown to play a significant role in comprehension and long-term storage of speech. Behavioral and neurophysiological studies suggest that filled pauses can help mitigate semantic and/or syntactic incongruity in spoken language. The purpose of the present study was to explore how filled pauses affect the processing of spontaneous speech in the listener. Brain activation of eight subjects was measured by electroencephalography (EEG), while they listened to recordings of Wizard-of-Oz travel booking dialogues.

The results show a P300 component in the Primary Motor Cortex, but not in the Broca or Wernicke areas. A possible interpretation could be that the listener is preparing to engage in speech. However, a larger sample is currently being collected.

Introduction

Spontaneous speech contains not only words with lexical meaning and/or grammatical function but also a considerable number of elements, commonly thought of as not part of the linguistic message. These elements include so-called disfluencies, some of which are filled pauses, repairs, repetitions, prolongations, truncations and unfilled pauses (Eklund, 2004). The term 'filled pause' is used to describe non-words like "uh" and "uhm", which are common in spontaneous speech. In fact they make up around 6% of words in spontaneous speech (Fox Tree, 1995; Eklund, 2004).

Corley & Hartsuiker (2003) also showed that filled pauses can increase listeners' attention and help them interpret the following utterance segment. Subjects were asked to press buttons according to instructions read out to them. When the name of the button was preceded by a filled pause, their response time was shorter than when it was not preceded by a filled pause.

Corley, MacGregor & Donaldson (2007) showed that the presence of filled pauses in utterances correlated with memory and perception improvement. In an event-related potential (ERP) study on memory, recordings of utterances with filled pauses before target words were played back to the subjects. Recordings of utterances with silent pauses were used as comparison. In a subsequent memory test subjects had to report whether target words, presented to them one at a time, had occurred during the previous session or not. The subjects were more successful in recognizing words preceded by filled pauses. EEG scans were performed starting at the onset of the target words. A clearly discernable N400 component was observed for semantically unpredictable words as opposed to predictable ones. This effect was significantly reduced when the words were preceded by filled pauses. These results suggest that filled pauses can affect how the listener processes spoken language and have long-term consequences for the representation of the message.

Osterhout & Holcomb (1992) reported from an EEG experiment where subjects were presented with written sentences containing either transitive or intransitive verbs. In some of the sentences manipulation produced a garden path sentence which elicited a P600 wave in the subjects, indicating that P600 is related to syntactic processing in the brain.

Kutas & Hillyard (1980) presented subjects with sentences manipulated according to degree of semantic congruity. Congruent sentences in which the final word was contextually predictable elicited different ERPs than incongruent sentences containing unpredictable final words. Sentences that were semantically incongruent elicited a clear N400 whereas congruent sentences did not.

We predicted that filled pauses evoke either an N400 or a P600 potential as shown in the studies above. This hypothesis has explanatory value for the mechanisms of the previously mentioned attention-enhancing function of filled pauses (Corley & Hartsuiker, 2003). Moreover, the present study is explorative in nature in that it uses spontaneous speech, in contrast to most previous EEG studies of speech perception.

Given the present knowledge of the effect of filled pauses on listeners' processing of subsequent utterance segments, it is clear that direct study of the immediate neurological reactions to filled pauses proper is of interest.

The aim of this study was to examine listeners' neural responses to filled pauses in Swedish speech. Cortical activity was recorded using EEG while the subjects listened to spontaneous speech in travel booking dialogs.

Method

Subjects

The study involved eight subjects (six men and two women) with a mean age of 39 years and an age range of 21 to 73 years. All subjects were native speakers of Swedish and reported typical hearing capacity. Six of the subjects were right-handed, while two considered themselves to be left-handed. Subjects were paid a small reward for their participation.

Apparatus

The cortical activation of the subjects was recorded using instruments from Electrical Geodesics Inc. (EGI), consisting of a Hydrocel GSN Sensor Net with 128 electrodes. These high impedance net types permit EEG measurement without requiring gel application which permits fast and convenient testing. The amplifier Net Amps 300 increased the signal of the high-impedance nets. To record and analyze the EEG data the EGI software Net Station 4.2 was used. The experiment was programmed in the Psychology Software Tools' software E-Prime 1.2.

Stimuli

The stimuli consisted of high-fidelity audio recordings from arranged phone calls to a travel booking service. The recordings were made in a "Wizard-of-Oz" setup using speakers (two males/two females) who were asked to make travel bookings according to instructions (see Eklund, 2004, section 3.4 for a detailed description of the data collection).

The dialogs were edited so that only the party booking the trip (customer/client) was heard and the responding party's (agent) speech was replaced with silence. The exact times for a total of 54 filled pauses of varying duration (200 to 1100 ms) were noted. Out of these, 37 were utterance-initial and 17 were utterance-medial. The times were used to manually identify corresponding sequences from the EEG scans which was necessary due to the nature of the stimuli. ERP data from a period of 1000 ms starting at stimulus onset were selected for analysis.

Procedure

The experiment was conducted in a sound attenuated, radio wave insulated and softly lit room with subjects seated in front of a monitor and a centrally positioned loud speaker. Subjects were asked to remain as still as possible, to blink as little as possible, and to keep their eyes fixed on the screen. The subjects were instructed to imagine that they were taking part in the conversation − assuming the role of the agent in the travel booking setting − but to remain silent. The total duration of the sound files was 11 min and 20 sec. The experimental session contained three short breaks, offering the subjects the opportunity to correct for any seating discomfort.

Processing of data

In order to analyze the EEG data for ERPs, several stages of data processing were required. A band pass filter was set to 0.3−30 Hz to remove body movement artefacts and eye blinks. A period of 100 ms immediately prior to stimulus onset was used as baseline. The data segments were then divided into three groups, each 1100 ms long, representing utterance-initial filled pauses, utterance-medial filled pauses and all filled pauses, respectively. Data with artefacts caused by bad electrode channels and muscle movements such as blinking were removed and omitted from analysis. Bad channels were then replaced with interpolated values from other electrodes in their vicinity. The cortex areas of interest roughly corresponded to Broca's area (electrodes 28, 34, 35, 36, 39, 40, 41, 42), Wernicke's area (electrodes 52, 53, 54, 59, 60, 61, 66, 67), and Primary Motor Cortex (electrodes 6, 7, 13, 29, 30, 31, 37, 55, 80, 87, 105, 106, 111, 112).The average voltage of each session was recorded and used as a subjective zero, as shown in Figure 1.

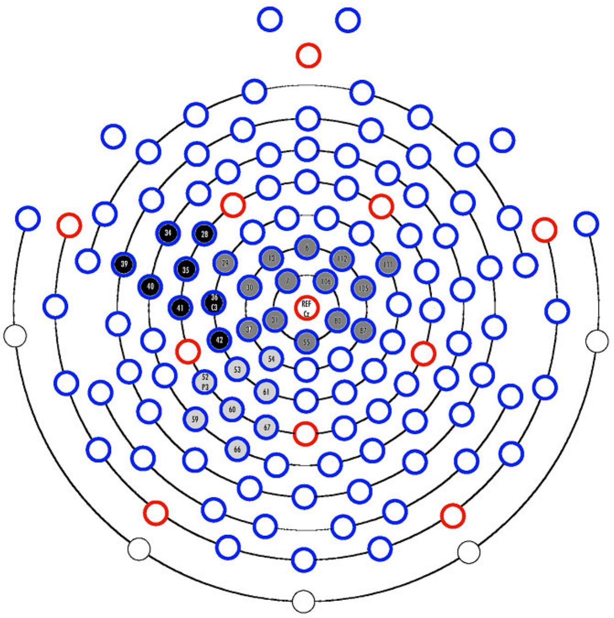

Figure 1. Sensor net viewed from above (nose up). Groups of electrodes roughly corresponding to Broca area marked in black, Wernicke area in bright grey and Primary Motor Cortex in dark grey.

Finally, nine average curves were calculated on the basis of selected electrode groups. The groups were selected according to their scalp localization and to visual data inspection.

Results

After visual inspection of the single electrode curves, only curves generated by the initial filled pauses (Figure 2) were selected for further analysis. No consistent pattern could be detected in the other groups. In addition to the electrode sites that can be expected to reflect language-related brain activation in Broca's and Wernicke's areas, a group of electrodes located at and around Primary Motor Cortex was also selected for analysis as a P300 component was observed in that area. The P300 peaked at 284 ms with 2.8 µV. The effect differed significantly ($p<.001$) from the baseline average. In the Wernicke area a later and weaker − however statistically significant ($p<.05$) − potential was observed than in Primary Motor Cortex.

At around 300 ms post-onset a negative tendency was visible, which turned to a very weak positive trend after approximately 800 ms. No clear trend for positive or negative voltage could be observed in the Broca area.

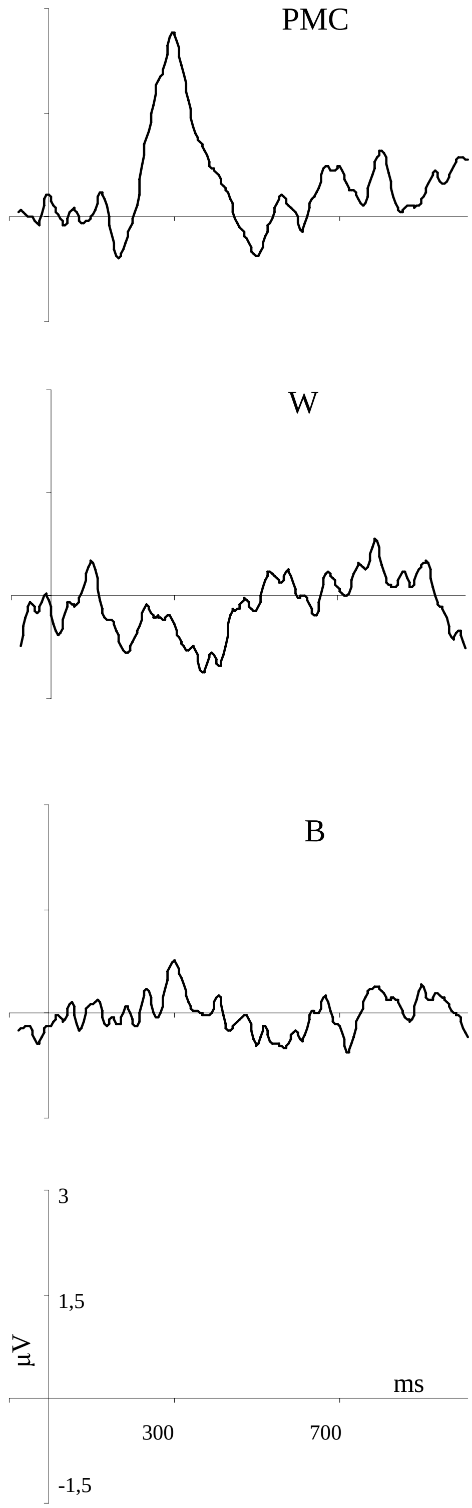

Figure 2. The average curves of initial filled pauses. Positive voltages in µV shown up. The intersection of the y-axis and the x-axis marks the baseline. A P300 component, peaking at 284 ms with 2.8 µV, is observed in Primary Motor Cortex (PMC). In Wernicke's area (W) the activation occurs somewhat later and the reaction is considerably weaker. In Broca's area (B) no consistent activation is observed.

Discussion

Contrary to our initial hypotheses, no palpable activation was observed in either Broca or Wernicke related areas.

However, the observed − and somewhat unexpected − effect in Primary Motor Cortex is no less interesting. The presence of a P300 component in or around the Primary Motor Cortex could suggest that the listener is preparing to engage in speech, and that filled pauses could act as a cue to the listener to initiate speech.

The fact that it can often be difficult to determine where the boundary between medial filled pauses and the rest of the utterance is could provide an explanation as to why it is difficult to discern distinct ERPs connected to medial filled pauses.

In contrast to the aforementioned study by Corley, MacGregor & Donaldson (2007) which used the word following the filled pause as ERP onset and examined whether the presence of a filled pause helped the listener or not, this study instead examined the immediate neurological response to the filled pause per se.

Previous research material has mainly consisted of laboratory speech. However, it is potentially problematic to use results from such studies to make far-reaching conclusions about processing of natural speech. In this respect the present study − using spontaneous speech – differs from past studies, although ecologically valid speech material makes it harder to control for confounding effects.

A larger participant number could result in more consistent results and a stronger effect.

Future studies could compare ERPs generated by utterance-initial filled pauses on the one hand and initial function words and/or initial content words on the other hand, as syntactically related function words and semantically related content words have been shown to generate different ERPs (Kutas & Hillyard, 1980; Luck, 2005; Osterhout & Holcomb, 1992). Results from such studies could provide information about how the brain deals with filled pauses in terms of semantics and syntax.

Acknowledgements

The data and basic experimental design used in this paper were provided by Robert Eklund as part of his post-doctoral fMRI project of disfluency perception at Karolinska Institute/ Stockholm Brain Institute, Dept of Clinical Neuroscience, MR-Center, with Professor Martin Ingvar as supervisor. This study was further funded by the Swedish Research Council (VR 421-2007-6400) and the Knut and Alice Wallenberg Foundation (KAW 2005.0115).

References

Corley, M. & Hartsuiker, R.J. (2003). Hesitation in speech can… um… help a listener understand. In *Proceedings of the 25th Meeting of the Cognitive Science Society*, 276−281.

Corley, M., MacGregor, L.J., & Donaldson, D.I. (2007). It’s the way that you, er, say it: Hesitations in speech affect language comprehension. *Cognition, 105*, 658−668.

Eklund, R. (2004). *Disfluency in Swedish human-human and human-machine travel booking dialogues*. PhD thesis, Linköping Studies in Science and Technology, Dissertation No. 882, Department of Computer and Information Science, Linköping University, Sweden.

Fox Tree, J.E. (1995). The effect of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language, 34*, 709−738.

Kutas, M. & Hillyard, S.A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*, 203−205.

Luck S.J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.

Osterhout, L. & Holcomb, P.J. (1992). Event-related potentials elicited by syntactic anomaly. *Journal of Memory and Language*, 31, 785−806.