ISCA Archive
http://www.isca-speech.org/archive

ITRW on Disfluency in
Spontaneous Speech (DiSS'01)
Edinburgh, Scotland, UK
August 29-31, 2001

Listeners' ERP Responses to False Starts and Repetitions in Spontaneous Speech

Jan McAllister, Susan Cato-Symonds and Blake Johnson

Department of Psychology
University of Auckland, New Zealand
j.mcallister@auckland.ac.nz

Abstract

Hindle [1] suggested that false starts and repetitions should be handled differently in a computational account of the processing of the two kinds of disfluency, and there is behavioural evidence that the human sentence processing mechanism likewise honours this distinction [2]. The same dichotomy was also evident in the electrophysiological data reported here. False starts and repetitions were identified in a corpus of spontaneous speech. Control items for the false starts were prepared by excising the reparanda to yield apparently fluent items. Continuous EEG was recorded while subjects listened to items containing the false starts, fluent false start controls, and first and second tokens of repetitions. Compared with identical words in their fluent controls, the false starts elicited a positive response similar to the P600 which is reported for syntactically anomalous words [3, 4, 5]. By contrast, second tokens of repetitions in general resulted in increased amplitude of the N400 [6]; yet, when the same repetitions were excised from context and presented list-fashion, they elicited the positive-going response which has been reported by other researchers [7].

1. Introduction

Although the speech we hear around us every day contains numerous disfluencies (interruptions to the smooth flow of speech), our conscious percept is that what we hear is, for the most part, grammatically well-formed. This suggests that our brains are organized in such a way that disfluencies are dealt with very rapidly and efficiently, and usually without conscious awareness. In this paper, we examine the brain responses that are elicited when a listener encounters a disfluency.

The following utterance contains two disfluencies:

> And he's got, he's got a little triangle hat and
> it – you can see a thin neck and a thick body.

The two disfluencies are, first, a repetition of the words "he's got", and second, a false start when the speaker abandons the original utterance at the word "it" and begins a fresh one with the words "you can see". Comprehension of these two sorts of disfluency, repetitions and false starts, was the focus of the study reported here.

When a speaker produces an utterance containing a disfluency such as "and it - you can see a thin neck and a thick body", it appears as though all or part of the original utterance "and it" is intended to be aborted and replaced with the new material. The term *reparandum* is used to refer to the material being aborted and *repair* to refer to material being substituted. Between the reparandum and repair the speaker may sometimes also insert material at the boundary between the reparandum and repair, such as a filled pause (e.g. "erm"), a silent pause or some other kind of phonological marker; such a marker is called the *editing term.*

One of the first researchers to attempt to model the comprehension of disfluent speech was Hindle [1]. A central feature of Hindle's model was the provision of distinct mechanisms for dealing with repetitions and false starts. Fox Tree [2] maintained this distinction between the two kinds of disfluency in the design of a set of monitoring experiments. She had subjects monitor for words that occurred soon after false starts or repetitions in spontaneous speech, and compared latencies with those for control materials in which the reparandum had been digitally excised. She found that monitoring times were slowed in items containing false starts, relative to their controls, suggesting that some processing cost was attached to the premature abandonment of an utterance and/or analysis of the repair. By contrast, subjects responded faster to targets that followed intact repetitions than to those in control materials without repetitions.

The goal in the present study was two-fold. First, we wished to establish whether disfluent utterances were distinguished from fluent counterparts by an electrophysiological marker. Second, we sought to determine whether Hindle's and Fox-Tree's distinction between repetitions and false starts was borne out in the data – specifically, whether different event-related potentials (ERPs) were elicited by the two kinds of disfluency.

In the last twenty years, ERP studies have provided many insights into language processing. The first report of a distinctly linguistic ERP was provided by Kutas and Hillyard [6] who found that semantically anomalous words were characterised by a negative-going wave, peaking at about 400 ms after stimulus onset, which they termed N400. The precise nature of N400 is still a matter of some debate. It is elicited in response to seemingly divergent stimuli, including line drawings, pronounceable non-words, low-frequency words and words in discourse that lacks cues to cohesion. However, the N400 is not elicited in response to syntactically anomalous material. Instead, syntactic anomaly elicits a positive-going response, termed the P600 [3, 4] or Syntactic Positive Shift [5], that may onset at 300 ms post-stimulus, or earlier in spoken language [4].

Deliberate (i.e., non-disfluent) repetitions are also associated with particular ERP effects. When repetitions occur in written or spoken lists, the response to the second token is more positive-going than that to the first token after 250 ms or 400 ms, respectively [7]. Repetitions that occur intentionally as part of connected discourse result in modulations to N400 and other components [8].

In the present study, false starts and repetitions were selected from an existing corpus of spontaneous speech. These disfluent materials, along with fluent control items that were created by digitally editing the false start materials, were played to subjects whose ERP responses to the same words in the fluent and disfluent contexts were recorded. In addition, the first and second tokens of the repetition items were excised from context and played list-fashion to the subjects, among sequences of non-repeated items, to determine whether the response to repetitions in context resembled that already reported for items repeated in lists.

2. Method

2.1. Materials

The digitized corpus of spontaneous speech from which the utterances were drawn is described by Murfitt and McAllister [9].

2.1.1. False starts

The false starts were selected first. Exclusion criteria included the presence of another disfluency in the immediate vicinity of the false start, overlapping talk by another speaker, overt editing terms at the boundary between the reparandum and repair, and word fragments in or immediately before the reparandum. One hundred and twenty eight false starts were initially identified as being potentially useable. These were excised, along with sufficient preceding and following context to permit listeners to perform the behavioural task (see below). Next, fluent control materials were made by editing copies of these experimental items so as to remove the reparandum. As previous researchers have noted, it is frequently possible to perform this editing in such a way that the resulting utterance is indistinguishable from unedited fluent speech [2]. To check the success of editing in the present materials, two volunteers checked the 128 edited items along with 20 fillers. Wherever either listener accurately identified splicing, the items were re-edited. The whole set of items was then checked by a further listener; any items where splicing was still detectable were discarded. Ninety-eight items survived this further test, and were included in the experiment. The false start materials represented the speech of nine speakers.

2.1.2. Repetitions in original context

The repetitions were next selected. The false start materials contained 17 repetitions that were not in the close vicinity of the false start itself (i.e. neither within, nor immediately adjacent to, the reparandum or the repair of a false start). From the speech of the nine speakers, a further 33 repetitions were identified, making 50 repetitions in all. The additional repetitions were taken from the digitised speech corpus along with sufficient context to make them comparable to the false start trials.

2.1.3. Tangrams

The forty tangram pictures that were used in Murfitt and McAllister's experiment [9] were digitised for use in the behavioural task.

Two experimental sequences were compiled. The same repetitions occurred in both sequences, but the false starts were divided into two sets. In one experimental sequence, half of the false starts were presented in their disfluent form while the other half were presented in their fluent form. In the second experimental sequence, items that were presented in their fluent form in the first sequence were now presented in their disfluent form, and vice versa.

Locations of the onsets of the target words were identified and these locations were used to position triggers that permitted segmentation of the continuous EEG record. The triggers were positioned at the start of the repair in the false start materials, at the start of the same word in the fluent controls, and at the start of the first and second tokens of repetitions.

Finally, utterances were paired with tangram pictures. Half the utterances were paired with a tangram picture that was consistent with the description given, and half with pictures that were inconsistent. Nine practice items and nine orientation trials were also prepared.

2.1.4. Repetitions in Lists

As noted in the introduction, well-established ERP responses are associated with the presentation of repeated items in list style, in both the auditory and visual modalities. An additional condition was therefore prepared. The 50 repetitions were excised from context and were arranged in lists along with 200 non-repeated filler items which were also drawn from the spontaneous corpus. Like the repetitions themselves, some of the non-repeated items were single words and some were multi-word sequences. Speech files were created, one for each speaker, in which the repeated and non-repeated items were presented, separated by one second of silence. One second of silence also separated the two tokens of any repeated items. As in the main experiment, the locations of the onsets of the each token of the repetitions was identified to allow placement of EEG segmentation triggers.

2.2. Subjects

Twenty members of the University of Auckland community took part in the experiment. They were paid NZ$20 for approximately 2 hours of their time. All subjects were right-handed, neurologically normal, and native speakers of English.

2.3. EEG

Electrical Geodesics 128 channel Ag/AgCl electrode nets were used. EEG was recorded continuously (250Hz sampling rate; 0.1-39.2Hz analogue band pass) during the experiment with Electrical Geodesics amplifiers (100 MΩ) and acquisition software running on a Power Macintosh 9600/200 computer with a National Instruments PC1-1200 12 bit analogue to digital conversion card. Electrode impedance ranged from 10 to 50 kΩ. EEG was initially acquired using a common vertex (Cz) reference, and subsequently re-referenced to averaged mastoids for purposes of analysis.

2.4. Procedure

Subjects were tested individually in a quiet room. They were instructed that on each trial they would hear a spoken description which would be immediately followed by a tangram picture. They were to decide whether the picture was consistent with the description that they had just heard. The speech materials were presented via earphones and the subjects saw the tangram pictures on a computer screen. Subjects initiated trials by pressing a computer key, which triggered the presentation of the speech material. While the spoken material was being presented, a fixation cross was shown in the centre of the screen, and subjects were asked to gaze at the cross while listening to the description. At the end of the spoken description, the tangram picture replaced the cross, and subjects gave their behavioural response via the computer keyboard.

After the main part of the experiment had been completed, subjects participated in the "repetition in lists" control condition. They were told that they were now going to hear lists of speech items, some of which consisted of several words, and some of which consisted of only one word. They were instructed that they should simply listen to the speech and silently identify the words that they were hearing.

3. Results

3.1. Behavioural task

The subjects made appropriate judgments about the tangrams 91% of the time, indicating that they were conscientiously attending to the meaning of the utterances.

3.2. ERPs

Due to space limitations, only results for midline electrodes are discussed here.

3.2.1. Responses to False Starts

Waveforms associated with disfluent and fluent items at electrode Pz are shown in Figure 1. Responses elicited by

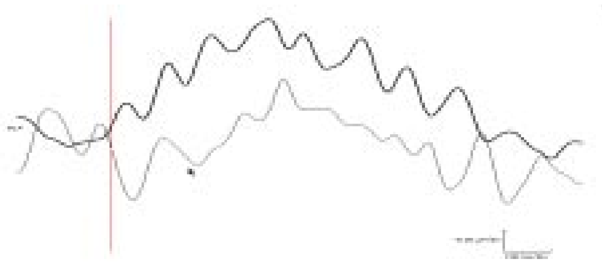

Figure 1: Averaged ERP response to false start stimuli; the heavy line shows responses to control (fluent) stimuli, the faint line, responses to experimental (disfluent) stimuli. The vertical bar indicates onset of the critical word.

disfluent stimuli, relative to their fluent controls, were characterised by a positive-going wave that was bilaterally distributed, and more pronounced at posterior sites. The positive response was similar to the P600 reported for continuous speech containing syntactic anomalies [4]. Statistical analysis of the ERP responses at the three midline sites Fz, Cz and Pz indicated a reliable early divergence of the waveforms for fluent and disfluent materials. In the 0-300 ms window, disfluent items were already eliciting a more positive-going response than fluent items ($F [1,19] = 5.87$, $p = 0.0255$); this effect was reliable only at electrodes Cz and Pz (Fluency X Electrode Site interaction: $F [2,38] = 4.24$, $p = 0.0218$). Electrode Site was not significant as a main effect ($F < 1$). In the 300-500 ms window, disfluent items were once again associated with more positive-going waveforms ($F [1,19] = 5.11$, $p = 0.0357$). The main effect of Electrode Site was also significant ($F [2,38] = 5.89$, $p = 0.0059$), with electrodes Cz and Pz more negative overall in this time window than Fz. The interaction between these variables was not statistically reliable ($F [2,38] = 2.11$, n.s.). In the 500-800 ms time window, a similar pattern emerged: fluency was once again significant ($F [1,19] = 9.45$, $p = 0.0062$), and electrode site was marginally so ($F [2,38] = 3.17$, $p = 0.0533$). There was no interaction between the variables ($F < 1$).

3.2.2. Repetitions in spontaneous speech

Figure 2: Averaged ERP response to repetition stimuli when these were heard in their original speech contexts. The heavy line shows responses to first tokens, the faint line, responses to second (i.e. disfluent) tokens. The vertical bar indicates onset of the critical word.

Figure 2 shows the responses at electrode Pz when repetitions were presented in their original (i.e. spontaneous speech) contexts. As figure 2 shows, responses to the first and second tokens elicited markedly different waveforms. The difference was apparent as an increased negativity associated with second (i.e., disfluent) tokens. This was bilaterally distributed in parietal and posterior regions.

In the 0-300 ms time window, neither the main effects nor the interaction were significant (all Fs approximately 1). In the 300-500 ms window, fluency was marginally significant (F [1,19] = 3.24, p=0.0879); neither the main effect of electrode site nor the fluency X electrode site interaction were significant (both Fs approximately 1). In the 500-800 ms window, neither main effects nor interaction were significant (all Fs approximately 1).

3.2.3. Repetitions in lists

Figure 3 shows ERP responses at electrode Pz when these same repetitions were presented in lists with 1 second of silence between first and second tokens. The lists also contained singleton filler items. Relative to the first token of the stimuli, repetitions were now characterised by greater positivity bilaterally at parietal and posterior sites in the latter part of the epoch. This positivity is broadly typical of items repeated in lists [7].

Figure 3: Averaged ERP response to first and second tokens of repetition stimuli when these were presented as items in a list. The heavy line represents the first tokens, the faint line the second tokens. The vertical bar indicates stimulus onset.

Consistent with other published studies, activity in two time windows, 0-400 ms and 400-800 ms, was analysed. The fluency effect shown in Figure 3 was significant in the earlier time window (F [1,19 = 4.85, p < 0.05), but neither the effect of electrode site nor the interaction were significant (both Fs < 1). In the second time window, there were differences between the three electrode sites overall (F [2, 38] = 6.20, p = 0.0047), and fluency (F [1, 19] = 5.94, p = 0.0248), but the two variables did not interact.

4. Discussion and Conclusion

The presence of the P600 effect associated with false starts indicates that listeners detect a syntactic anomaly when processing these items. The earliness of the divergence between the disfluent and fluent false start materials suggests that the anomaly may actually be associated with an event at the end of the reparandum, rather than the beginning of the repair. None of the experimental items contained overt editing terms such as “erm” at the boundary between the reparandum and the repair, but it is possible that other (e.g. prosodic) cues may have indicated that the item was about to become disfluent. Second tokens of repetitions, when they were heard in their original sentence contexts, gave rise to a negative response, relative to the corresponding first tokens. This finding indicates that false starts and are differently processed, as anticipated by Hindle [1] and Fox Tree [2]. It is possible that the repetition response found here may be a variant of the discourse repetition response reported by van Petten et al. [8]. Interestingly, the present result is also consistent with the ERP response to repetition blindness [10]. Repetition blindness/deafness is one mechanism that has been suggested to account for listeners’ frequent failure to retain a conscious percept of the occurrence of a disfluent repetition [11].

5. References

[1] Hindle, D. ``Deterministic parsing of syntactic non-fluencies ", *Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics,* 123-128, 1983.

[2] Fox Tree, J. (1995). “The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech”, *Journal of Memory and Language 34, 1995, p 709-738.*

[3] Osterhout, L., and Holcomb, P. “Event-related brain potentials elicited by syntactic anomaly”, *Journal of Memory and Language 31, 1992, p 785-806.*

[4] Osterhout, L., and Holcomb, P. “Event-related potentials and syntactic anomaly: Evidence of anomaly detection during the perception of continuous speech”, *Language and Cognitive Processes 8, 1993, p 413-437.*

[5] Hagoort, P., Brown, C., and Groothusen, J. “The Syntactic Positive Shist (SPS) as an ERP-measure of syntactic processing”, *Language and Cognitive Processes 8, 1993, p 439-484.*

[6] Kutas, M., and Hillyard, S. “Reading senseless sentences: Brain potentials reflect semantic incongruity”, *Science 207, 3, 1980, p 203-205.*

[7] Rugg, M., Doyle, M.C., and Melan, C. “An event-related potential study of the effects of within- and across-modality word repetition”, *Language and Cognitive Processes 8, 1993, p 357-378.*

[8] Van Petten, C., Kutas, M., Kluender, R., Mitchener, M., et al. “Fractionating the word-repetition effect with event-related potentials”, *Journal of Cognitive Neuroscience 3, 2, 1991, p 131-150.*

[9] Murfitt, T., and McAllister, J. “The effect of production variables in monologue and dialogue on comprehension by novel listeners”, *Language and Speech, in press.*

[10] Schendan, H, Kanwisher, N, and Kutas, M. “Early brain potentials link repetition blindness, priming and novelty detection.” *Neuroreport: Neuroscience. Vol 8(8), May 1997, 1943-1948.*

[11] Bard, E.G., and Lickley, R. “Disfluency Deafness: Graceful Failure in the Recognition of Running Speech.”, Proceedings of CogSci '98.