Neuropsychologia 48 (2010) 3982–3992

Contents lists available at ScienceDirect

Neuropsychologia

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

Listening to the sound of silence: disfluent silent pauses in speech have consequences for listeners

Lucy J. MacGregor[a,*,1], Martin Corley[b], David I. Donaldson[c]

[a] Institute of Psychological Sciences, University of Leeds, Leeds, United Kingdom
[b] Psychology, PPLS, University of Edinburgh, United Kingdom
[c] Department of Psychology, University of Stirling, United Kingdom

ARTICLE INFO

Article history:
Received 28 December 2009
Received in revised form 9 June 2010
Accepted 23 September 2010
Available online 13 October 2010

Keywords:
Language comprehension
Disfluency
ERPs
Recognition memory
N400
PMN
LPC

ABSTRACT

Silent pauses are a common form of disfluency in speech yet little attention has been paid to them in the psycholinguistic literature. The present paper investigates the consequences of such silences for listeners, using an Event-Related Potential (ERP) paradigm. Participants heard utterances ending in predictable or unpredictable words, some of which included a disfluent silence before the target. In common with previous findings using *er* disfluencies, the N400 difference between predictable and unpredictable words was attenuated for the utterances that included silent pauses, suggesting a reduction in the relative processing benefit for predictable words. An earlier relative negativity, topographically distinct from the N400 effect and identifiable as a Phonological Mismatch Negativity (PMN), was found for fluent utterances only. This suggests that only in the fluent condition did participants perceive the phonology of unpredictable words to mismatch with their expectations. By contrast, for disfluent utterances only, unpredictable words gave rise to a late left frontal positivity, an effect previously observed following *ers* and disfluent repetitions. We suggest that this effect reflects the engagement of working memory processes that occurs when fluent speech is resumed. Using a surprise recognition memory test, we also show that listeners were more likely to recognise words which had been encountered after silent pauses, demonstrating that silence affects not only the process of language comprehension but also its eventual outcome. We argue that, from a listener's perspective, one critical feature of disfluency is the temporal delay which it adds to the speech signal.

1. Introduction

Spoken language is rarely continuously fluent. As well as producing the *ums*, *ers*, repetitions, restarts and repairs that occur up to six times per hundred words of speech (Bortfeld, Leon, Bloom, Schober, & Brennan, 2001; Fox Tree, 1995), speakers are often silent mid-utterance. Silences can be deliberate: for example, speakers may use silence as a rhetorical device, or to maintain the prosodic structure of an utterance. Equally, however, silences can reflect linguistic performance factors such as difficulty in planning or retrieving upcoming words (Goldman-Eisler, 1958a, 1958b; Kircher, Brammer, Levelt, Bartels, & McGuire, 2004; Maclay & Osgood, 1959; Martin, 1967). Given their myriad possible causes (see also Duez, 1985; Ferreira, 2007; Zellner, 1994), different types of silences can be difficult to distinguish, particularly when they occur between clauses. For this reason researchers investigating the imperfections of speech have typically ignored interruptions that result in a silent pause (Bortfeld et al., 2001), or have conflated them with filled pauses like *er* and *um* (e.g., Hawkins, 1971). By contrast, in the present paper we focus explicitly on silent pauses, examining the ways in which they affect listeners' processing of speech, and their subsequent representations of utterances. We use a design that is directly comparable to those of those of two previous studies (Corley, MacGregor, & Donaldson, 2007; MacGregor, Corley, & Donaldson, 2009), allowing us to compare the effects of silences to those of other disfluencies.

A recent body of research has shown that mid-utterance disruptions to fluent speech do have consequences for listeners. To date, however, the majority of studies have focused on the filled pause *er*, which is typically associated with production difficulties. A range of methodologies have been used to show that *ers* can affect language processing in different ways. Studies measuring eye movements have shown that following a disfluent pause there is an increase in the probability of an initial eye movement to a discourse-new (Arnold, Tanenhaus, Altmann, & Fagnano, 2004) or unfamiliar item (Arnold, Hudson Kam, & Tanenhaus, 2007) from a constrained set of referents. From these results it has been argued that disfluent

* Corresponding author at: MRC Cognition and Brain Sciences Unit, 15 Chaucer Rd, Cambridge, CB2 7EF, United Kingdom.
E-mail address: lucy.macgregor@mrc-cbu.cam.ac.uk (L.J. MacGregor).

0028-3932/$ – see front matter
doi:10.1016/j.neuropsychologia.2010.09.024

pauses can increase listeners' expectations for the mention of a lexical item that is more difficult for the speaker to say. Consistent with such an interpretation is evidence from Event-Related Potentials (ERPs), which shows that an *er* can also affect the ease with which subsequent predictable compared to unpredictable words are integrated into their contexts (Corley et al., 2007). In addition to the effects on the immediate process of comprehension, the disfluent pause *er* can also have longer-lasting effects: most notably, words heard following an *er* are more likely to be remembered during a surprise later recognition memory test (Collard, Corley, MacGregor, & Donaldson, 2008; Corley et al., 2007).

There is limited evidence regarding the effects of other types of disfluencies, but some research suggests that not all disfluencies affect listeners equally. Although repetitions typically occur in similar situations to *er*s and may reflect similar difficulties for the speaker, they tend to have different consequences for listeners: *er* and *oh* have a facilitative effect when participants are asked to monitor for subsequent words (Fox Tree, 2001; Fox Tree & Schrock, 1999), whereas repetitions appear to have little effect on processing (Fox Tree, 1995; MacGregor et al., 2009). Interestingly, however, there is evidence to suggest that repetitions and *er*s may entail the engagement of similar post-disfluency processes that occur as listeners resume fluent processing after an interruption (MacGregor et al., 2009).

Silent pauses present a different challenge to comprehension than other disfluencies. Listeners encountering silent pauses are not faced by the introduction of new phonetic or lexical material (Clark & Fox Tree, 2002, suggest that *um* and *uh* are words which mark the speaker's difficulty in continuing). On the other hand, disfluent silences may occur in similar circumstances to filled pauses, disrupting the temporal flow of speech, and delaying the onset of subsequent information. Perhaps for this reason, several studies investigating the effects of an *er* have used silent pauses as a baseline condition. When response times to targets in an object selection task are measured, silence appears to give rise to a similar facilitation effect to that associated with an *er* (Brennan & Schober, 2001). One interpretation of these data is that the effect is due to the temporal delay the disruption introduces into the utterance, a suggestion that receives support from evidence that *er*s and environmentally plausible interruptions (such as dog barks) have similar effects on listeners' final interpretations of syntactically ambiguous sentences (Bailey & Ferreira, 2003). However, an explicit comparison of the disfluency *er* with silence suggests that the two disfluencies may not give rise to identical effects: response times to targets in a word monitoring task are faster following an *er* than following a silent pause (Fox Tree, 2001, although it should be pointed out that the durations of the interruptions were not matched in this study).

Other studies have used fully fluent utterances as the baseline with which to compare the processing of disfluent utterances. A number of these studies have made use of ERPs—measures of electrical brain activity recorded (as EEGs) from electrodes placed on the human scalp, time-locked to the onset of a cognitive event of interest and averaged over multiple events. ERPs provide an index of neural activity that reveals the time course of cognitive processing; the very precise temporal resolution of ERPs makes them a particularly useful tool for monitoring listeners' cognitive processing of speech. In the first ERP study of disfluency Corley et al. examined listeners' responses to utterances containing an *er* whilst measuring the N400 effect. The N400 (Kutas & Hillyard, 1980, 1984) has been widely used in studies of language processing because it provides an index of the ease with which the meaning of a word can be accessed and integrated into its context (see Kutas, Van Petten, & Kluender, 2006). In Corley et al.'s study, participants heard utterances ending either in predictable (high cloze) or unpredictable (low cloze) words. Critically, half of the experimental materials included *er* disfluencies immediately preceding the targets. For fluent utterances, unpredictable words resulted in greater centro-parietal negativity than predictable words, maximal around 400 ms, interpretable as a standard N400 effect. However, this effect was greatly attenuated in the disfluent condition, suggesting that there was little difference in integration difficulty for unpredictable compared to predictable words following a disfluency. One likely explanation is that disfluency affected the extent to which upcoming words could be predicted. A subsequent study (MacGregor et al., 2009) investigated repetition disfluencies, in which the word prior to the target word was repeated. Using a similar design to that of Corley et al. (2007), no attenuation of the N400 was found in disfluent conditions. However, in this case unpredictable targets in disfluent utterances gave rise to a late left frontal positivity, an effect which was also observed following *er*s. MacGregor et al. (2009) suggested that despite their differences (on the N400 effect), both *er*s and repetitions interrupted listeners' fluent comprehension processes. According to the account proposed, the resumption of fluent comprehension engaged memory control processes associated with retrieval of the preceding context or updating of working memory.

To our knowledge, only one ERP study has explicitly compared fluent speech to speech containing between-word silences. Besson, Faita, Czternasty, and Kutas (1997) asked participants to listen to utterances that were either highly constrained proverbs ending in predictable words, or unconstrained utterances ending in unpredictable words. Sometimes the utterance-final critical word was delayed unexpectedly by 600 ms of silence. The unexpected silent pause elicited a negative–positive complex: the negative component peaked around 100 ms after pause onset and was followed by positive component which peaked around 350–400 ms. These components, particularly the positivity, were larger when the pause followed a highly constrained utterance (a proverb) than when it followed a weakly constrained utterance. A similar N1–P2 complex has also been observed when a pause appeared within (rather than between) words and listeners had to explicitly detect the presence of the pauses (Mattys, Pleydell-Pearce, Melhorn, & Whitecross, 2005), and when a pause was present in the context of a musical phrase rather than in connected speech (Besson & Faita, 1995; Besson, Faita, & Requin, 1994). The N1–P2 complex has been interpreted as reflecting temporal disruption (Besson et al., 1997; Mattys et al., 2005), although a simpler interpretation based on the acoustic deviance of silent pauses has not been ruled out.

The presence of a clear ERP response to unexpected delays is not in itself particularly surprising. Of greater interest, and more relevant to the current study, is that the design of Besson et al. study also enabled an assessment of the impact of the interruption on the processing of subsequent predictable compared to unpredictable words, through the observation of its impact on the N400 effect. Besson et al. (1997) showed that for both fluent utterances and utterances containing an interruption, unpredictable words elicited an N400 relative to predictable words, indicating an increase in the difficulty with which unpredictable words could be processed. However, the N400 effects were not identical; there were observable differences in the timings. For fluent utterances the N400 onset around 150 ms whereas the onset was delayed by around 250 ms following an interruption. The authors suggested that the later onset of the N400 following an unexpected pause may reflect the absence of co-articulatory cues (which provide listeners with early information about the identity of the upcoming word), or the surprise of not hearing a word when it was expected.

Taken together, the evidence concerning the effects of silence as compared to other disfluencies is currently equivocal. Support for the possibility that silences are similar or dissimilar to other disfluencies can be found where behavioural methods rely on subsidiary tasks (e.g., Brennan & Schober, 2001; Fox Tree, 2001). Where