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

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

[a] *MRC Cognition and Brain Sciences Unit, Cambridge*
[b] *Psychology, PPLS, University of Edinburgh*
[c] *Department of Psychology, University of Stirling*

---

**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 *er*s 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

*Corresponding author. Address: MRC Cognition and Brain Sciences Unit, 15 Chaucer Rd, Cambridge UK.
*Email address:* `lucy.macgregor@mrc-cbu.cam.ac.uk` (Lucy J. MacGregor)

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.

*Key words:* Language comprehension, Disfluency, ERPs, Recognition memory, N400, PMN, LPC

---

## 1. Introduction

Spoken language is rarely continuously fluent. As well as producing the *um*s, *er*s, repetitions, restarts and repairs that occur up to six times per hundred words of speech (Bortfeld, Leon, Bloom, Schober, and 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,b; Kircher, Brammer, Levelt, Bartels, and McGuire, 2004; Maclay and 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, and Donaldson, 2007; MacGregor, Corley, and 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 *er*s 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, and Fagnano, 2004) or unfamiliar item (Arnold, Hudson Kam, and Tanenhaus, 2007) from a constrained set of referents. From these results it has been argued that disfluent 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 (Corley et al., 2007; Collard, Corley, MacGregor, and Donaldson, 2008).

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 and 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 and 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 and 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 and 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 and 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, and 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 400ms, 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 600ms of silence. The unexpected silent pause elicited a negative-positive complex: the negative component peaked around 100ms after pause onset and was followed by positive component which peaked around 350–400ms. 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, and Whitecross, 2005), and when a pause was present in the context of a musical phrase rather than in connected speech (Besson and Faita, 1995; Besson, Faita, and 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 150ms whereas the onset was delayed by around 250ms 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 and Schober, 2001; Fox Tree, 2001). Where online ERP measures are used, it has been shown that silence before a word has an effect on timing (Besson et al., 1997), but this finding is not easily comparable to other work which has focused on amplitude variations due to disfluencies such as *er* or repetitions. In the present paper we attempt to resolve this issue, reporting an experiment which investigates the impact of disfluent silent pauses in spoken utterances on listeners, using a design and materials based on those reported in Corley et al. (2007) and MacGregor et al. (2009). Specifically, to assess language processing online we use ERPs; as a starting point we consider whether silent pauses affect the ease with which subsequent words can be integrated into their contexts, as indexed by the amplitude of the N400 effect. Additionally, we investigate the later representation of the spoken utterances using a visual old-new recognition memory paradigm, in which participants are asked to identify which of a series of words were presented previously.

## 2. Method

### *2.1. Participants*

Sixteen native British English speakers (6 male; mean age 22 years; range 17–29 years; all right-handed) who reported no hearing or reading difficulties, and had no known neurological impairment, participated for financial compensation (£5 per hour) or course credit. Informed consent was obtained in accordance with the University of Stirling Psychology Ethics Committee guidelines. Six participants (2 males) were excluded from ERP analyses due to artefacts in the EEG recordings. Behavioural data was recorded and analysed from all sixteen participants.

### *2.2. Materials*

The stimuli were 160 highly constrained fluent and disfluent utterances ending in predictable (cloze probability 0.84, range 0.52–1) or unpredictable (cloze probability 0) target words and were based on those used in Corley et al. (2007). Utterances were constructed in pairs, such that each predictable word also served as an unpredictable word for a corresponding utterance. Furthermore, predictable and unpredictable targets completed fluent and disfluent utterances so that, across participants, each target appeared in every condition (Latin square design). This double counterbalancing ensured that targets were perfectly controlled for grammatical class, duration, frequency, imageability, and concreteness and meant that each participant heard all sentence frames and target words once only.

In an effort to avoid potential smearing of ERP effect onsets (e.g., Van Petten, Coulson, Rubin, Plante, and Parks, 1999), the predictable and unpredictable targets within each utterance pair had different phonetic onsets so that the targets would be acoustically distinguishable at word onset.[1] Utterances were selected from a larger set which had been submitted to a cloze probability pre-test on the World Wide Web (`www.language-experiments.org`) using a minimum of 17 participants per sentence. Table 1 illustrates an example material set.

Table 1: Example stimulus set comprising two highly constraining sentence frames, crossed with two target words, which were either predictable or unpredictable in context. Target words are shown in bold. Half of the utterances included an interruption in the form of a silent pause before the target word, which is indicated by (. . . ).

| | | | |
|---|---|---|---|
| Predictable | Everyone's got bad habits and mine is biting my | (. . . ) | **nails** |
| | That drink's too hot; I've just burnt my | (. . . ) | **tongue** |
| Unpredictable | Everyone's got bad habits and mine is biting my | (. . . ) | **tongue** |
| | That drink's too hot; I've just burnt my | (. . . ) | **nails** |

Fluent and disfluent versions of the sentence frames were recorded at a natural speaking rate by a female native English speaker who had no speech production difficulties. For each utterance, the final target word was replaced by the pseudotarget word 'pen' which meant that there were no acoustic cues to the upcoming word. Any prosodic cues to an upcoming /p/ were constant across conditions. Disfluent contexts were originally recorded with an *er* be-

[1] One target pair inadvertently had the same acoustic onset.

fore the target and with other features of disfluency that often co-occur with silent pauses, to enhance the ecological validity of the stimuli as in previous work investigating the effects of the disfluent pause *er* (Corley et al., 2007). These additional features, such as vowel lengthening, were elicited where the speaker felt it to be natural and resulted in lengthened pre-target words in 35% of disfluent utterances and overall a longer mean duration of disfluent (3910 ms) compared to fluent (2702 ms) sentence frames. Following recording, utterances were edited (using Adobe Audition, `www.adobe.com`) to excise the *er*, so that a silent pause of identical duration remained. Pseudo-targets were excised and replaced by target words which had been recorded as utterance-final words in separate carrier sentences. Targets were spliced onto the fluent and disfluent contexts such that acoustically identical tokens appeared across each condition (mean duration of the target word was 450ms).

An additional 80 filler utterances of varying constraint were included to mask the nature of the experimental manipulation. Forty were fluent and 40 contained disfluencies of various types (repetitions, *er*s, silent pauses, and repairs) in various locations. Before presentation, all stimuli were converted to 16-bit 22050 Hz .wav files, and their amplitudes were normalised so that the acoustic volume was approximately matched across stimuli. Four versions of the experiment were created, for counterbalancing purposes, each containing 160 experimental utterances (40 each of fluent predictable, fluent unpredictable, disfluent predictable, and disfluent unpredictable) together with the 80 filler utterances.

### *2.3. Procedure*

There were two parts to the experiment, which lasted around 1 hour after setup (approximately 1 hour). In the first part, participants were told that they would hear a series of utterances which were re-recorded excerpts from natural conversations. Participants were further advised that because the utterances would be heard out of context, some would make more sense than others. They were instructed to listen for understanding, just as they would in a natural situation. There was no other task. To minimise the introduction of artifacts into the EEG recording, it was emphasised to participants that they should relax, keep as still as possible, and fixate their eyes on a cross in the centre of the screen.

One hundred and sixty experimental utterances were presented auditorily, in a random order, interspersed with fillers. Utterances were presented in two blocks lasting approximately 15 minutes each, separated by a break of a few minutes. The start of each utterance was indicated visually (for 250ms) by a yellow fixation cross on a black screen, which flashed blue once (for 250ms) and returned to yellow as the utterance began. The fixation cross remained on the screen for the duration of the utterance to discourage eye movements. After each utterance the screen was blanked (for 1500ms).

After the first part of the experiment, participants took part in a surprise recognition memory test for the utterance final "old" words. These words had been either contextually predictable or unpredictable, and had been heard in either fluent or disfluent utterances. They were interspersed with 160 frequency-matched "new" foils, which had not been heard at any point in the first part of the experiment. Targets were presented visually, and participants discriminated between old and new words as accurately as possible by pressing one of two response keys with index fingers (counterbalanced across participants). The start of each presentation was indicated by the appearance of a fixation cross (for 400ms), which was replaced by the target word (for 750ms), after which the screen was blanked (for 1750ms).

### *2.4. ERP recording and pre-processing*

Electrophysiological data was recorded (Neuroscan 4.2 Acquire software, `www.neuro.com`), processed (Neuroscan 4.3 Edit software, `www.neuro.com`) and analysed in the Psychological Imaging Laboratory at the University of Stirling (`www.pil.stir.ac.uk`) using methods which are standard in the cognitive electrophysiology field. EEGs were recorded from 61 Ag/AgCl scalp electrodes embedded in an elasticated cap, based on an extended version of the international 10-20 system (Jasper, 1958). Data were recorded using a left mastoid reference, and re-referenced offline to the average of left and right mastoid recordings. Electro-oculograms (EOGs) were recorded to monitor vertical and horizontal eye movements. Electrode impedances were kept below 5k$\Omega$. The analogue EEG and EOG recordings were amplified (band pass filter 0.01–40Hz), and continuously digitised (16 bit) at a sampling frequency of 200Hz.

Before offline averaging, the continuous EEG files for each participant were segmented into 1350ms epochs starting 150ms before the critical words, baseline corrected using the 150ms pre-target interval and screened for artifacts. Epochs were excluded when any channel became saturated (exceeding 495 $\mu$V), when drift (absolute difference in amplitude between the first and last data point of each individual epoch) was greater than 33.75 $\mu$V, or when amplitude on any channel (excluding VEOG) was greater than 75 $\mu$V. A minimum of 16 artefact-free trials was required from each participant, in each condition, to ensure an acceptable signal-to-noise ratio. The screening process resulted in the loss of 40% of the trials, predominantly due to drift, with no difference between conditions. The effect of eye-blink artifacts was minimised by estimating and correcting their contribution to the ERP waveforms using a regression procedure: for each participant, an average blink was created from 32 blinks and the contribution of the blink was removed from all other channels on a point-by-point basis. Waveforms were baseline corrected by subtracting the mean amplitude over the interval preceding the critical word and smoothed over 5 points so that each sampling point represents the average over the two previous and two subsequent points.

Grand average ERPs were formed time-locked to the critical words in each condition (with mean number of trials per condition given in parentheses): fluent predictable (24), fluent unpredictable (23), disfluent predictable (24), disfluent unpredictable (25).

### *2.5. Data analysis*

ERPs were quantified by measuring the mean amplitude over three time windows of interest: 300–500ms and 600–900ms based on previous research on the effects of disfluency on language processing (Corley et al., 2007; MacGregor et al., 2009); 50–200ms based on observations of the waveforms. Differences between conditions were assessed using analyses of variance (ANOVAs) with an alpha level of .05. The main analyses had factors of Fluency [fluent, disfluent], Predictability [predictable, unpredictable], Location [F, FC, C, CP, P], Hemisphere [left, right], and Site [superior: electrode 1/2, medial: electrode 3/4, inferior: electrode 5/6]. Figure 3 shows the electrodes used in the analysis. Significant interactions were explored further with the appropriate subsidiary ANOVAs.

Differences in the scalp distributions of significant effects (unpredictable minus predictable) between conditions (fluent versus disfluent and across time windows) were assessed after normalisation for amplitude differences at all 61 electrodes using the Max/Min method (McCarthy and Wood, 1985). All analyses made use of Greenhouse-Geisser corrections where appropriate, and are reported using corrected $F$ and $p$ values.

## 3. Results

### *3.1. ERP Results*

Figures 1 and 2 show ERPs time-locked to the utterance-final word onsets for fluent and disfluent utterances respectively. For fluent utterances, unpredictable words lead to a greater negativity over the 300–500ms time window relative to predictable words. A similar pattern is observed for utterances which included a disfluent silent pause. For fluent utterances, the effect is broadly distributed over the scalp, but appears larger over central/centro-parietal and midline sites, typical of an N400 effect. For disfluent utterances, the effect is less broadly distributed over the scalp but still appears larger at posterior locations, again typical of an N400 effect.

It is clear from observation of the waveforms that differences between fluent and disfluent utterances are also apparent before and after the N400 time window. Preceding the N400 effect, fluent utterances show a relative negativity for unpredictable words that is focused over frontal sites to a greater extent than is typical of N400 effects. By contrast, unpredictable words in disfluent utterances show a weak posterior negativity accompanied by some frontal positivity. Beyond the standard N400 time window, unpredictable words in fluent utterances continue to elicit relative negativity, although this appears reduced in both amplitude and spread compared to the N400 effect. For disfluent utterances, unpredictable words are associated with relative positivity, particularly over the left hemisphere, accompanied by a weak posterior negativity. The topography of the ERP effects (unpredictable minus predictable) for fluent and disfluent utterances in three time windows of interest, are depicted in Figure 4.

#### *3.1.1. 300–500ms*

The main analysis (see Table 2) demonstrated greater negativity for unpredictable compared to predictable words, particularly over posterior and superior electrodes (predictability*site, $p = .019$; predictability*location*site, $p = .043$). Importantly, the N400 effect was statistically larger and more widespread for the fluent (mean amplitude over parietal electrodes = $1.341\mu$V) than for the disfluent ($0.719\mu$V) condition. The characteristic distribution of the N400 effect and amplitude difference between fluent and disfluent conditions can be clearly seen in the middle panel of Figure 4.

For the fluent condition only, the N400 was particularly prevalent over superior sites (predictability*site, $p = .003$) and showed a trend towards being larger at more posterior locations (predictability*location, $p = .053$). By contrast, the analysis for the disfluent condition failed to show any significant effects, reflecting the absence of a reliable N400 predictability effect in this case.

#### *3.1.2. 50–200ms*

The main analysis (Table 3 indicated a number of differences between fluent and disfluent conditions (see Figure 4). Unpredictable words elicited a relative negativity in the fluent condition, but a relative positivity over the left hemisphere in the disfluent condition (fluency*predictability*hemisphere, $p = .012$). The negativity in the fluent condition was larger at frontal locations

Table 2: Analyses of Variance (ANOVAs) on mean ERP amplitudes in the 300–500ms time window. Factor labels are as follows: Flu = Fluency, Pre = Predictability, Loc = Location, Hem = Hemisphere, Sit = Site

| ANOVA | *df* | *F* | *MSE* | $\eta_p^2$ | *p* |
|---|---|---|---|---|---|
| Flu X Pre X Loc X Hem X Sit | | | | | |
| fluency*predictability*site | 2,18 | 10.41 | 1.95 | .54 | .007 |
| predictability*site | 2,18 | 6.15 | 3.06 | .41 | .019 |
| predictability*location*site | 8,72 | 3.54 | 1.12 | .28 | .043 |
| Fluent: Pre X Loc X Hem X Sit | | | | | |
| predictability*site | 2,18 | 11.43 | 3.34 | .56 | .003 |
| predictability | 1,9 | 4.21 | 95.86 | .32 | .071 |
| predictability*location | 4,36 | 4.45 | 14.67 | .33 | .053 |

over superior sites in contrast to the frontal positivity in the disfluent condition (fluency*predictability*site, $p = .050$; fluency*predictability*location*site, $p = .047$). Across both fluent and disfluent conditions, there was a relative negativity towards a posterior location (predictability*site, $p = .043$) and towards superior sites (predictability*site, $p = .018$).

Statistics for the fluent condition alone supported the presence of a frontal negativity over superior sites (predictability*site, $p = .004$; predictability*location*site, $p = .002$). For the disfluent condition alone, there was a gradient from relative positivity on the left to negativity on the right (predictability*hemisphere, $p = .028$; predictability*hemisphere*site, $p = .041$). Analysis of the normalised predictability effects supported the presence of distributional differences between the fluent and disfluent conditions (fluency*hemisphere, $p = .003$; fluency*hemisphere*site, $p = .007$).

### *3.1.3. 600–900ms*

The main analysis (Table 4, see Figure 4) indicated the presence of a relative positivity for unpredictable words at frontal locations for the disfluent condition,

Table 3: Analyses of Variance (ANOVAs) on mean ERP amplitudes in the 50–200ms time window. Factor labels are as follows: Flu = Fluency, Pre = Predictability, Loc = Location, Hem = Hemisphere, Sit = Site

| ANOVA | *df* | *F* | *MSE* | $\eta_p^2$ | *p* |
|---|---|---|---|---|---|
| Flu X Pre X Loc X Hem X Sit | | | | | |
| fluency*predictability*site | 2,18 | 4.80 | 8.77 | .35 | .05 |
| fluency*predictability*hemisphere | 1,9 | 9.75 | 18.83 | .52 | .012 |
| fluency*predictability*location*site | 8,72 | 3.2 | 1.34 | .26 | .047 |
| predictability*location | 4,36 | 4.53 | 21.37 | .34 | .043 |
| predictability*site | 2,18 | 7.74 | 9.69 | .46 | .018 |
| Fluent: Pre X Loc X Hem X Sit | | | | | |
| predictability*site | 1,9 | 13.57 | 1.38 | .60 | .004 |
| predictability*location*site | 2,18 | 6.01 | 0.22 | .40 | .002 |
| Disfluent: Pre X Loc X Hem X Sit | | | | | |
| predictability*hemisphere | 1,9 | 6.85 | 1.77 | .43 | .028 |
| predictability*hemisphere*site | 2,18 | 5.13 | 0.22 | .36 | .041 |
| Normalised Predictability Effect: Flu X Loc X Hem X Sit | | | | | |
| fluency*hemisphere | 1,9 | 16.02 | 0.38 | .640 | .003 |
| fluency*hemisphere*site | 2,18 | 9.28 | 0.08 | .51 | .007 |

which was not apparent for the fluent condition (fluency*predictability*location, $p$ = .053). There was a relative negativity at superior sites that was largest over central locations (predictability*site, $p = .028$; predictability*location*site, $p < .001$), and a positivity over the left hemisphere that was greater at sites away from the midline (predictability* hemisphere*site, $p = .048$).

For the fluent condition alone, statistics indicated a central-focussed negativity over superior sites (predictability*site, $p = .036$; predictability*location*site, $p = .010$). By contrast, for the disfluent condition there was a relative positivity at frontal locations and over the left hemisphere (predictability*location,

$p$ = .001; predictability*hemisphere, $p$ = .023). Three-way interactions reflected the greater positivity at frontal superior sites and at centro-parietal inferior sites compared to negativity over posterior sites (predictability*location* site, $p$ = .001), and the focus of the positivity over the left hemisphere, particularly at inferior sites (predictability*hemisphere*site, $p$ = .008). Analysis of the normalised predictability effects also reflected the presence of a frontal positivity for the disfluent condition compared to the central negativity for the fluent condition (fluency*location, $p$ = .087).

Table 4: Analyses of Variance (ANOVAs) on mean ERP amplitudes in the 600–900ms time window.

| ANOVA | *df* | $F$ | $MSE$ | $\eta_p^2$ | $p$ |
|---|---|---|---|---|---|
| Flu X Pre X Loc X Hem X Sit | | | | | |
| fluency*predictability*location | 4,36 | 4.18 | 9.56 | .32 | .053 |
| predictability*site | 2,18 | 6.24 | 1.84 | .41 | .028 |
| predictability*location*site | 8,72 | 15.38 | 0.44 | .63 | $< .001$ |
| predictability*hemisphere*site | 2,18 | 4.83 | 1.12 | .35 | .048 |
| Fluent: Pre X Loc X Hem X Sit | | | | | |
| predictability*site | 2,18 | 5.40 | 3.59 | .38 | .036 |
| predictability*location*site | 8,72 | 4.81 | 0.45 | .35 | .010 |
| Disfluent: Pre X Loc X Hem X Sit | | | | | |
| predictability*location | 4,36 | 10.80 | 3.05 | .55 | .001 |
| predictability*hemisphere | 1,9 | 7.45 | 9.21 | .45 | .023 |
| predictability*location*site | 8,72 | 8.21 | 0.86 | .48 | .001 |
| predictability*hemisphere*site | 2,18 | 9.41 | 0.61 | .51 | .008 |
| Normalised Predictability Effect: Flu X Loc X Hem X Sit | | | | | |
| fluency*location | 4,60 | 3.30 | 1.87 | .27 | .087 |

#### *3.1.4. Effects over time*

Unpredictable items in fluent utterances gave rise to significant negativity at all three time windows, therefore further analyses of the fluent ERPs were performed on the normalised data, to assess whether the distributions of the predictability effects changed over time. A comparison of the effects between the 50–200 ms and 300–500 ms time windows indicated that the frontal negativity observed over midline sites during the early time window was more focused than in the later time window where the distribution was broader and maximal at the posterior location (epoch*location*site: $F(8, 72) = 7.556$, $MSE = 0.071$, $\eta_p^2 = .456$, $p = .002$). By contrast, there were no significant topographic differences between the effects observed in the 300–500ms and 600–900ms time windows, suggesting that the later effect reflects a continuation of the N400 effect. Effects of predictability were not observed across successive time windows for disfluent items and therefore topographic comparisons were not made for this condition.

### *3.2. Recognition memory results*

Memory performance was quantified as the proportion of occasions on which all 16 participants correctly recognised words as "old" words, separated as a function of fluency and predictability. Overall, 52% of the old words were correctly recognised.[2] Figure 5 shows the recognition probability of utterance-final words by fluency and predictability.

Analysis of categorical data using ANOVA violates the ANOVA assumptions, which can lead to spurious results, even after transformation of the proportional data (see Jaeger, 2008). Moreover, ANOVA analyses cannot take random variance due to participants and items into account simultaneously. We therefore used mixed effects logit modelling, which additionally allowed us to simultaneously account for by-participant and by-item variances. A model with factors

[2]One item was corrupted and excluded from the analyses. In addition, 5% of the items were not responded to by participants (within the allocated time) resulting in no data for these items. These items are excluded from the analyses.

of fluency, predictability, and random variance was not improved by the addition of an interaction term ($\chi^2(1) = .236$, $p < 1$). As is shown in Figure 5, participants were 1.32 times as likely to recognise words which had previously been preceded by a silence ($z = 2.26$, $p = 0.02$), and they were also more likely to recognise words which were unpredictable in context (OR = 1.48, $z = 3.16$, $p < 0.01$).

## 4. Discussion

The present study provides evidence that silence has consequences for listeners' processing and retention of the linguistic material it interrupts. Participants listened to utterances which were either fluent or contained a silent pause that rendered them disfluent. Most notably, the N400 effect associated with unpredictable compared to predictable words in fluent utterances was attenuated when a silent pause preceded the target word. We were also able to detect an early negativity associated with unpredictable targets in fluent items, which differed topographically from the later N400 effect. Furthermore, we observed a late left frontal positivity associated with unpredictable items which followed a silence. After listening to the utterances, participants took part in a recognition memory test for words they had heard. Participants were more likely to remember words which had followed silent pauses as well as words that were contextually unpredictable.

### *4.1. Silent pauses affect semantic integration*

When participants listen to recorded utterances, they find it more difficult to integrate unpredictable compared to predictable words into their contexts, as indexed by the N400 effect (see, e.g., Brown and Hagoort, 1993; Hagoort, 2008; Hagoort and Brown, 2000). The results presented here demonstrate that when the target words are preceded by a disfluent silence, the N400 effect, and thus the difference in integration difficulty, is reduced.

The difference in pre-target baselines between fluent and disfluent conditions (word versus silence) means that in the present experimental design, direct comparisons of the effects of particular types of (predictable or unpredictable) target between conditions are not warranted. Therefore, it is not possible to definitively attribute the attenuation of the N400 effect to a change to the processing of either predictable or unpredictable words. However, it seems unlikely that a disfluency would facilitate the ease of integrating a subsequent contextually unpredictable word, particularly when there was no limited set of referents to restrict what the speaker might mention. It is worth noting that facilitated processing of an unpredictable entity might be plausible in the visual world paradigm where there is an array of possible referents because in this case, specific predictions for a contextually unpredictable referent could be formed (c.f. Arnold et al., 2004, 2007). Instead, we suggest that disfluency increases the integration difficulty of subsequent words, which will be particularly apparent for predictable words, resulting in a decrease in the amplitude of the N400 effect.

Very similar consequences for semantic integration are observed when participants hear utterances including *er* disfluencies (Corley et al., 2007); however, when the disfluencies consist instead of repetitions, there is no attenuation of the N400 effect (MacGregor et al., 2009). When considered alongside *er*s and disfluent repetitions, the effects of silent pauses support a refined version of the suggestion that, from the listeners' perspective, a critical feature of disfluency is the temporal delay which it adds to the speech signal (e.g., Brennan and Schober, 2001; Bailey and Ferreira, 2003). What appears to be crucial is that there is an interruption between the critical word and the preceding context into which it is integrated. Where a repetition of the last part of the context occurs, there is no such interruption and effects on semantic integration and recognition memory are not observed. Although the evidence to date is consistent with a purely delay account of disfluency processing, it is important to note that temporal delay could be viewed as a mere side-effect of an acoustic mismatch. Whereas repetitions can only be distinguished from an utterance's message by context, silences and filled pauses are acoustically distinct from the words surrounding them. Of course, accounts which focus on either delay or acoustic mismatch introduced by a disfluency are not mutually exclusive; it seems likely that a combination of factors is at play. Whichever account turns out to be correct, a number of questions remain outstanding. For example, it is unknown what the minimal duration of a disfluent delay is, although there is evidence showing that a delay is perceived to be shorter when it is filled with an *er* than when it is silent (Brennan and Williams, 1995). Similarly, the effects of an interruption on comprehension may be all-or-nothing, or graded, depending on timing. Finally, it is also the case that the disfluent utterances included features of disfluency such as lengthening, before the interruption itself, and therefore that the effects cannot be conclusively attributed to the presence of a silent pause alone; future work should disentangle the contributing factors.

*4.2. Silent pauses prevent the early detection of phonological mismatches*

The present study showed that unpredictable words in fluent utterances gave rise to a relative negativity which onset earlier than a typical N400 effect. Somewhat similarly, Besson et al. (1997) reported a reliable negativity for low- compared to high-cloze target words in connected speech, which onset around 150ms earlier in unmanipulated utterances than in a condition where the targets were preceded by a 600ms silence. Besson et al. interpreted the negativities in both conditions as N400 effects which varied in timing due to co-articulatory cues present only in the fluent utterances. Importantly, however, they did not directly compare the amplitudes nor distributions of the N400 effects between conditions. In the present study, we were able to establish that the early negativity for fluent utterances had a fronto-central distribution that was topographically distinct from the N400 effect (Besson et al. used only 7 electrodes whereas we recorded data from 61). We also note that, in the present study the fluent stimuli did not include co-articulatory cues to the targets, due to the way in which they were constructed (see section 2.2) and therefore the early effect in the present experiment can be associated with the target word onset. In fact the distribution of the effect is compatible with its identification as a Phonological Mismatch Negativity (PMN: Connolly, Stewart, and Phillips, 1990; Connolly and Phillips, 1994). The PMN has been previously been observed in N400 paradigms and is thought to reflect detection of a mismatch of the phonological onset of the presented word with a listener's expectations.

For disfluent utterances there was no indication of a PMN, suggesting that there was no clash of the phonology with participants' expectations. We did find two interactions (involving predictability and hemisphere) over the same time window for disfluent items, but the topography and antecedent conditions do not suggest a specific interpretation, and we therefore do not discuss this effect further. In terms of accounting for the absence of the PMN in disfluent conditions, one explanation is that the temporal delay introduced by the disfluent silence simply rendered any phonological mismatch less salient. Another possibility is that the delay disrupted normal linguistic processing, such as the generation of predictions about the phonology of upcoming words. Although the latter interpretation is preliminary, it is consistent with the finding that disfluencies (notably *er*) can affect listeners' predictions under some circumstances (Arnold et al., 2004, 2007), and with the view, expressed above, that in the present study the N400 effect indexes the extent to which listeners form specific predictions.

### *4.3. Silent pauses have consequences when processing is resumed*

The presence of a silent pause in speech affects linguistic processing as indexed by changes to the N400 and PMN ERP effects. In addition, we also observed a third ERP difference: for disfluent utterances only, unpredictable words were associated with a Late Positive Complex (LPC) at left frontal locations over the 600–900ms time window. The effect is similar in timing and distribution to positivities observed in response to unpredictable words following both *er*s and disfluent repetitions (MacGregor et al., 2009), and similar to positivities which are sometimes observed in conditions associated with N400 elicitation (for example, Federmeier, Wlotko, De Ochoa-Dewald, and Kutas, 2007, reported a similar positivity when unexpected words followed highly constraining written contexts).

A number of interpretations of late positivities in language studies have been proposed, including the suggestion that it reflects difficulties with the processing and integration of multiple types of linguistic information, both structural and conceptual in nature (Kaan, Harris, Gibson, and Holcomb, 2000; Münte, Heinze, Matzke, Wieringa, and Johannes, 1998; Friederici, 2002; Eckstein and Friederici, 2005; Kuperberg, Caplan, Sitnikova, Eddy, and Holcomb, 2006). However, it should be noted that across experiments, the distributions of reported late positivities differ and are therefore likely to reflect the engagement of a number of different processes.

The frontal distribution of the positivity observed in the present study is similar to positivities that are observed in studies of memory and are associated with retrieval effort (Ranganath and Paller, 1999; Rugg, Allan, and Birch, 2000), and the distribution is consistent with a semantic memory-related generator in the left inferior prefrontal cortex (Coulson and Wu, 2005). Thus functional interpretations of LPCs observed during language processing are sometimes related to aspects of memory control. Consistent with a memory account of the LPC, memory is likely to play a role in cases where listeners must resume structural and semantic interpretation of an utterance following an interruption, whether it be a silence, an *er*, or a repetition. In the present study, the unpredictable words shared fewer semantic associations with the preceding context than the predictable words. We suggest that LPCs observed following disfluency are associated with the updating of working memory (cf. Van Petten, Kutas, Kluender, Mitchiner, and McIsaac, 1991), which may be more difficult to achieve when the target word is unpredictable (and therefore not a very effective cue for the preceding context).

### *4.4. Silent pauses enhance memory*

A large body of research has demonstrated clear associations between the cognitive processes engaged during encoding of various stimuli and the likelihood that the information is remembered later (the DM, or difference in memory effect, e.g., Besson and Kutas, 1993; Craik and Lockhart, 1972; Paller, Kutas, and Mayes, 1987; Rugg, Mark, Walla, Schloerscheidt, Birch, and Allan, 1998; Wagner, Schacter, Rotte, Koutstaal, Maril, Dale, Rosen, and Buckner, 1998). The finding reported in the present paper that participants are more likely to later recognise words which have been preceded by a silent pause sits well within this body of literature in demonstrating that the impact of silence on immediate linguistic processing has consequences for the later representation of what was heard. The present study was not designed for the analysis of ERPs contingent on subsequent memory performance and therefore there were insufficient trial numbers to consider DM effects.

One way in which silence may affect linguistic encoding is through attentional capture. Several studies have reported that an *er* can heighten listeners' attention during the processing of speech (Brennan and Schober, 2001; Collard et al., 2008; Fox Tree, 2001). Because listeners are attending to the words that occur immediately after the disfluency, this can account for the finding that words heard following an *er* are more likely to be later recognised (Corley et al., 2007; Collard et al., 2008). The memory advantage for words following silence may well be because silences capture attention in much the same way as *er*s.

We suggest that once attention has been captured, the listener may be able to determine that the next words the speaker utters may not be what was initially predicted. In contexts where there are limited numbers of candidate referents they may then be able to predict likely alternatives (Arnold et al., 2007); a more general strategy may be to abandon prediction altogether. Either of these accounts would explain the observed attenuation of the N400 effect because the differences in predictability (and hence integrability) between high- and low-cloze items would be reduced; abandoning prediction would also implicate attention, because top-down processes would no longer be able to compensate for bottom-up information.

Across studies with different disfluencies the N400 and recognition memory results pattern together (*er*s and silence reduce the N400 effect and enhance recognition memory: Corley et al. (2007) whereas disfluent repetitions have no effect on either: MacGregor et al. (2009)), which raises interesting questions about the relationship between linguistic processes indexed by the N400 and memory encoding processes. Although more evidence is clearly required, it seems possible that the N400 provides an index of at least one element of linguistic processing that leads to later recognition. From a memory encoding perspective, an increase in the cognitive processing needed to achieve semantic integration during language comprehension is analogous to an increase in 'depth of processing' (Craik and Lockhart, 1972). Deeper encoding results in a stronger, more elaborate memory trace and thus the item is more likely to be later remembered. As discussed earlier, we believe that the attenuation of the N400 effect following disfluent pauses is driven largely by an increase in integration difficulty for the predictable words. The increase in memorability for words heard following a silence (and *er*s) may, at least in part, reflect deeper processing of difficult-to-integrate words.

Depth of processing is thought to produce selective increases in recollection-based memory processes (Rugg et al., 1998) rather than familiarity-based memory. We therefore suggest that the increase in memorability for words preceded by disfluent pauses, which were more difficult to integrate, rests largely on recollection. Similarly, we would expect recollection to support the enhanced recognition observed for unpredictable relative to predictable words in fluent utterances. To our knowledge, only one study has focused specifically on linguistic encoding (integration) processes indexed by the N400 and its relationship to later recognition memory. This study (Meyer, Mecklinger, and Friederici, 2007) reported a correlation between the amplitude of the N400 during comprehension and the size of the familiarity-related ERP old/new effect during recognition. However, there was no difference in the probability of correctly remembering words as a function of the N400 amplitude. Clearly, more studies are needed to determine whether or not there is a relationship between the N400 effect, encoding, and later recognition memory and its sensitivity to different contextual variables.

### *4.5. Silent pauses should not be considered equivalent to* ers

According to the measures employed in the current experiment, silent pause disfluencies affect listeners' language processing and recognition memory in ways similar to those those observed previously with *er*s. However, it would be wrong to conclude from this that silence and *er*s serve entirely interchangeable functions in human discourse. Investigations of listeners' assessments of speakers have suggested that filled and silent hesitations are distinct: participants are likely to judge speakers as being more confident in their answers to general knowledge questions when a hesitation is silent than when a filled pause is used (Brennan and Williams, 1995), but the use of silent as opposed to filled pauses makes speakers seem less relaxed (Christenfeld, 1995). Filled pauses and silences may also serve different functions from the speaker's perspective. For example, the vocalisation of an *er* means it can be used by speakers to hold the floor (Clark and Fox Tree, 2002), whereas a silence offers other speakers the opportunity to easily interject. A full account of the effects that disfluencies have on listeners must therefore integrate evidence concerning metalinguistic judgements with the online evidence reported here, to explain the ways in which disfluencies affect communication at all levels of processing.

## 5. Conclusions

Silent pauses have similar effects on listeners' language processing and memory as those observed previously with *er*s, but largely different effects to those observed with disfluent repetitions. As indexed by the ERP N400 effect, disfluent silences affect the ease with which listeners can integrate words into their contexts. Moreover, the absence of a PMN in disfluent contexts is compatible with the suggestion that integration difficulty is affected because listeners abandon predictions of upcoming content when speakers are disfluent. Whether prediction is affected by a disfluency or not, when fluent speech is resumed, listeners must update their working memory by re-activating the preceding context for further processing . Re-activation may be more difficult when the target word is unpredictable and is therefore a poor cue for the preceding context. Consistent with this view, the present experiment demonstrates that, as in previous experiments with both disfluent repetitions and *er*s, unpredictable words in disfluent contexts resulted in a late left frontal positivity. These changes to online speech comprehension processes have longer-term consequences as revealed by enhanced recognition memory for target words that had been heard following a silent pause.

Taken together, our findings suggest that silence in speech has consequences for the way in which an utterance is processed and for the representation of what is said. In common with all disfluencies, silent pauses interrupt the ongoing process of comprehension; the ERP data presented here reveal how, in real time, listeners are able to flexibly re-engage in linguistic processing, allowing them to successfully understand the message their interlocutors are attempting to convey.

**Acknowledgements**

Thanks to Catriona Bruce for help with data collection and to two reviewers for their helpful comments. The research was partially funded by an ESRC postgraduate studentship (LJM). DID is a member of the SINAPSE Collaboration (`www.sinapse.ac.uk`), a Pooling Initiative funded by the Scottish Funding Council and the Chief Scientist Office of the Scottish Executive.

Figure 1: ERPs for fluent utterances relative to predictable (solid lines) or unpredictable (dotted lines) target word onsets. Positive is plotted up. The central column represents the midline sites (from top: frontal (F), fronto-central (FC), central (C), centro-parietal (CP), parietal (P), occipito-parietal (PO)); the left-hand and right-hand columns represent averages of three electrodes to the left or right of the midline respectively, as used in analyses.

Figure 2: ERPs for disfluent utterances including a silent pause relative to predictable (solid lines) or unpredictable (dotted lines) target word onsets. Positive is plotted up. The central column represents the midline sites (from top: frontal (F), fronto-central (FC), central (C), centro-parietal (CP), parietal (P), occipito-parietal (PO)); the left-hand and right-hand columns represent averages of three electrodes to the left or right of the midline respectively, as used in analyses.

Figure 3: Schematic maps of the 61 electrodes sites with sites used in the analyses highlighted. Electrode Cz is labelled for reference.

Figure 4: Scalp topographies showing the effects (unpredictable minus predictable) over three time windows: 50–200ms, 300–500ms and 600-900ms, for (a) *fluent* and (b) *disfluent* utterances.

Figure 5: Proportion of correctly recognised words that were originally predictable (black) or unpredictable (grey) in their contexts, for fluent and disfluent utterances.

**References**

Arnold, J. E., Hudson Kam, C. L., Tanenhaus, M. K., 2007. If you say thee uh you are describing something hard: The on-line attribution of disfluency during reference resolution. Journal of Experimental Psychology: Learning, Memory and Cognition 33, 914–930.

Arnold, J. E., Tanenhaus, M. K., Altmann, R. J., Fagnano, M., 2004. The old and thee, uh, new. disfluency and reference resolution. Psychological Science 15, 578–582.

Bailey, K. G. D., Ferreira, F., 2003. Disfluencies affect the parsing of garden-path sentences. Journal of Memory and Language 49, 183–200.

Besson, M., Faita, F., 1995. An event-related potential (ERP) study of musical expectancy: Comparison of musicians with nonmusicians. Journal of Experimental Psychology: Human Perception and Performance 21, 1278–1296.

Besson, M., Kutas, M., 1993. The many facets of repetition: A cued-recall and event-related potential analysis of repeating words in the same versus different sentence contexts. Journal of Experimental Psychology: Learning, Memory and Cognition 19, 1115–1133.

Besson, M., Faita, F., Czternasty, C., Kutas, M., 1997. What's in a pause: Event-related potential analysis of temporal disruptions in written and spoken sentences. Biological Psychology 46, 3–23.

Besson, M., Faita, F., Requin, J., 1994. Brain waves associated with musical incongruities differ for musicians and non-musicians. Neuroscience Letters 168, 101–105.

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., Brennan, S. E., 2001. Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. Language and Speech 44, 123–147.

Brennan, S. E., Schober, M. F., 2001. How listeners compensate for disfluencies in spontaneous speech. Journal of Memory and Language 44, 274–296.

Brennan, S. E., Williams, M., 1995. The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive state of speakers. Journal of Memory and Language 34, 383–398.

Brown, C. M., Hagoort, P., 1993. The processing nature of the N400: Evidence from masked priming. Journal of Cognitive Neuroscience 5, 34–44.

Christenfeld, N., 1995. Does it hurt to say um? Journal of Nonverbal Behaviour 19, 171–186.

Clark, H. H., Fox Tree, J. E., 2002. Using uh and um in spontaneous speaking. Cognition 84, 73–111.

Collard, P., Corley, M., MacGregor, L. J., Donaldson, D. I., 2008. Attention orienting effects of hesitations in speech: Evidence from ERPs. Journal of Experimental Psychology: Learning, Memory and Cognition 34, 696–702.

Connolly, J. F., Phillips, N. A., 1994. Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. Journal of Cognitive Neuroscience 6, 256–266.

Connolly, J. F., Stewart, S. H., Phillips, N. A., 1990. The effects of processing requirements on neurophysiological responses to spoken sentences. Brain and Language 39, 302–318.

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.

Coulson, S., Wu, Y. C., 2005. Right hemisphere activation of joke-related information: An event-related brain potential study. Journal of Cognitive Neuroscience 17, 494–506.

Craik, F. I. M., Lockhart, R. S., 1972. Levels of processing: A framework for memory research. Journal of Verbal Learning and Verbal Behaviour 11, 671–684.

Duez, D., 1985. Perception of silent pauses in continuous speech. Language and Speech 28, 377–389.

Eckstein, K., Friederici, A. D., 2005. Late interaction of syntactic and prosodic processes in sentence comprehension as revealed by ERPs. Cognitive Brain Research 25, 130-143.

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., Kutas, M., 2007. Multiple effects of sentential constraint on word processing. Brain Research 1146, 75–84.

Ferreira, F., 2007. Prosody and performance in language production. Language and Cognitive Processes 22, 1151–1177.

Fox Tree, J. E., 1995. The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. Journal of Memory and Language 34, 709–738.

Fox Tree, J. E., 2001. Listeners' uses of um and uh in speech comprehension. Memory and Cognition 29, 320–326.

Fox Tree, J. E., Schrock, J. C., 1999. Discourse markers in spontaneous speech: Oh what a difference an oh makes. Journal of Memory and Language 40, 280–295.

Friederici, A.D.,2002. Towards a neural basis for auditory sentence processing. Trends in Cognitive Science 6, 78–84.

Goldman-Eisler, F., 1958a. The predictability of words in context and the length of pauses in speech. Language and Speech 1, 226–231.

Goldman-Eisler, F., 1958b. Speech production and the predictability of words in context. Quarterly Journal of Experimental Psychology 10, 96–106.

Hagoort, P., 2008. The fractionation of spoken language understanding by measuring electrical and magnetic brain signals. Philosophical Transactions of the Royal Society of London B: Biological Sciences 363, 1055–1069.

Hagoort, P., Brown, C. M., 2000. ERP effects of listening to speech: Semantic ERP effects. Neuropsychologia 38, 1518–1530.

Hawkins, R. R., 1971. The syntactic location of hesitation pauses. Language and Speech 14, 277–288.

Jaeger, F. T., 2008. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. Journal of Memory and Language 59, 434–446.

Jasper, H. H., 1958. Report to the committee on methods of clinical examination in electroencephalography. Appendix: The ten-twenty system of the international federation. Electroencephalography and Clinical Neurophysiology 10, 370–375.

Kaan, E., Harris, T., Gibson, E., Holcomb, P. J. 2000. The P600 as an index of syntactic integration difficulty. Language and Cognitive Processes 15, 159–201.

Kircher, T. J., Brammer, M. J., Levelt, W. J. M., Bartels, M., McGuire, P. K., 2004. Pausing for thought: Engagement of left temporal cortex during pauses in speech. NeuroImage 21, 84–90.

Kuperberg, G. R., Caplan, D., Sitnikova, T., Eddy, M., Holcomb, P. J. Neural correlates of processing syntactic, semantic and thematic relationships in sentences. Language and Cognitive Processes 21, 489–530.

Kutas, M., Hillyard, S. A., 1980. Reading senseless sentences: Brain potentials reflect semantic incongruity. Science 207, 203–205.

Kutas, M., Hillyard, S. A., 1984. Brain potentials during reading reflect word expectancy and semantic association. Nature 307, 161–163.

Kutas, M., Van Petten, C. K., Kluender, R., 2006. Psycholinguistics electrified II (1994-2005). In: Traxler, M., Gernsbacher, M. A. (Eds.), Handbook of Psycholinguistics, 2nd Edition. Elsevier Press, San Diego, CA, pp. 83–143.

MacGregor, L. J., Corley, M., Donaldson, D. I., 2009. Not all disfluencies are are equal: The effects of disfluent repetitions on language comprehension. Brain and Language 111, 36–45.

Maclay, H., Osgood, C. E., 1959. Hesitation phenomena in English speech. Word 15, 19–44.

Magne, C., Astésano, Aramaki, M., Ystad, S., Kronland-Martinet, R., Besson, M., 2007. Influence of syllabic lengthening on semantic processing in spoken French: Behavioral and electrophysiological evidence. Cerebral Cortex 17, 2659–2668.

Martin, J. G., 1967. Hesitations in the speaker's production and listener's reproduction of utterances. Journal of Verbal Learning and Verbal Behavior 6, 903–909.

Mattys, S. L., Pleydell-Pearce, C. W., Melhorn, J. F., Whitecross, S. E., 2005. Detecting silent pauses in speech. A new tool for measuring on-line lexical and semantic processing. Psychological Science 16, 958–964.

McCarthy, G., Wood, C. C., 1985. Scalp distributions of event-related potentials: An ambiguity associated with analysis of variance models. Electroencephalography and Clinical Neurophysiology 62, 203–208.

Meyer, P., Mecklinger, A., Friederici, A. D., 2007. Bridging the gap between the semantic N400 and the early old/new memory effect. Neuroreport 18, 1009–1013.

Münte, T. F., Heinze, H. J. Matzke, M. Wieringa, B. M., Johannes, S., 1998. Brain potentials and syntactic violations revisited: No evidence for specificity of the syntactic positive shift. Neuropsychologia 36, 217–226.

Paller, K. A., Kutas, M., Mayes, A. R., 1987. Neural correlates of encoding in an incidental learning paradigm. Electroencephalography and Clinical Neurophsysiology 67, 360–371.

Ranganath, C., Paller, K. A., 1999. Frontal brain potentials during recognition are modulated by requirements to retrieve perceptual detail. Neuron 22, 605–613.

Rugg, M. D., Allan, K., Birch, C. S., 2000. Electrophysiological evidence for the modulation of retrieval orientation by depth of study processing. Journal of Cognitive Neuroscience 12, 664–678.

Rugg, M. D., Mark, R. E., Walla, P., Schloerscheidt, A. M., Birch, C. S., Allan, K., 1998. Dissociation of the neural correlates of implicit and explicit memory. Nature 392, 595–598.

Urbach, T. P., Kutas, M. 2006. Interpreting event-related brain potential (ERP) distributions: Implications of baseline potentials and variability with application to amplitude normalization by vector scaling. Biological Psychology 72, 333–343.

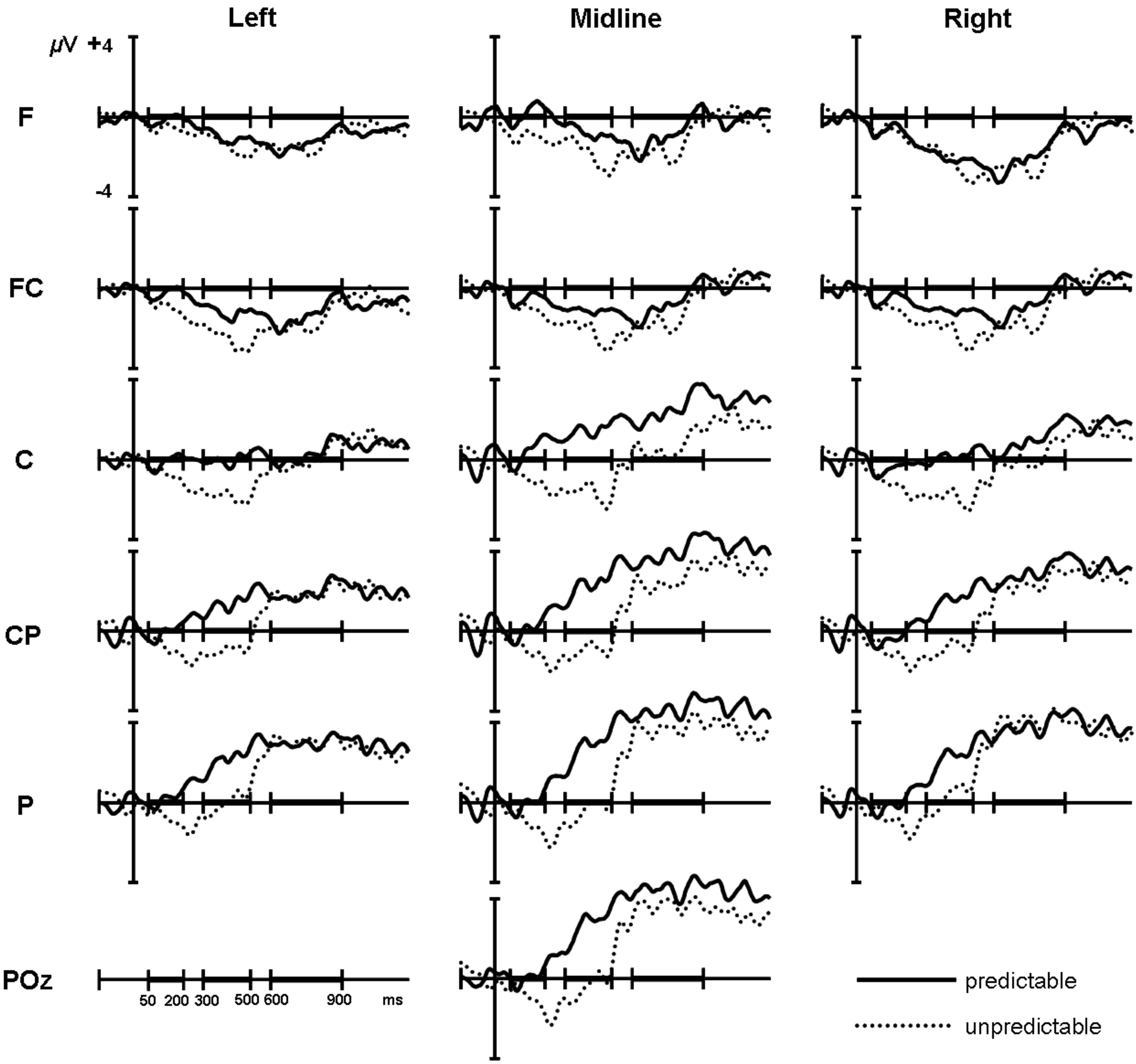
Van Petten, C. M., Coulson, S., Rubin, S., Plante, E., Parks, M., 1999. Time course of word identification and semantic integration in spoken language. Journal of Experimental Psychology: Learning, Memory, and Cognition 25, 394–417.

Van Petten, C. M., Kutas, M., Kluender, R., Mitchiner, M., McIsaac, H., 1991. Fractionating the word repetition effect with event-related potentials. Journal of Cognitive Neuroscience 3, 131–150.

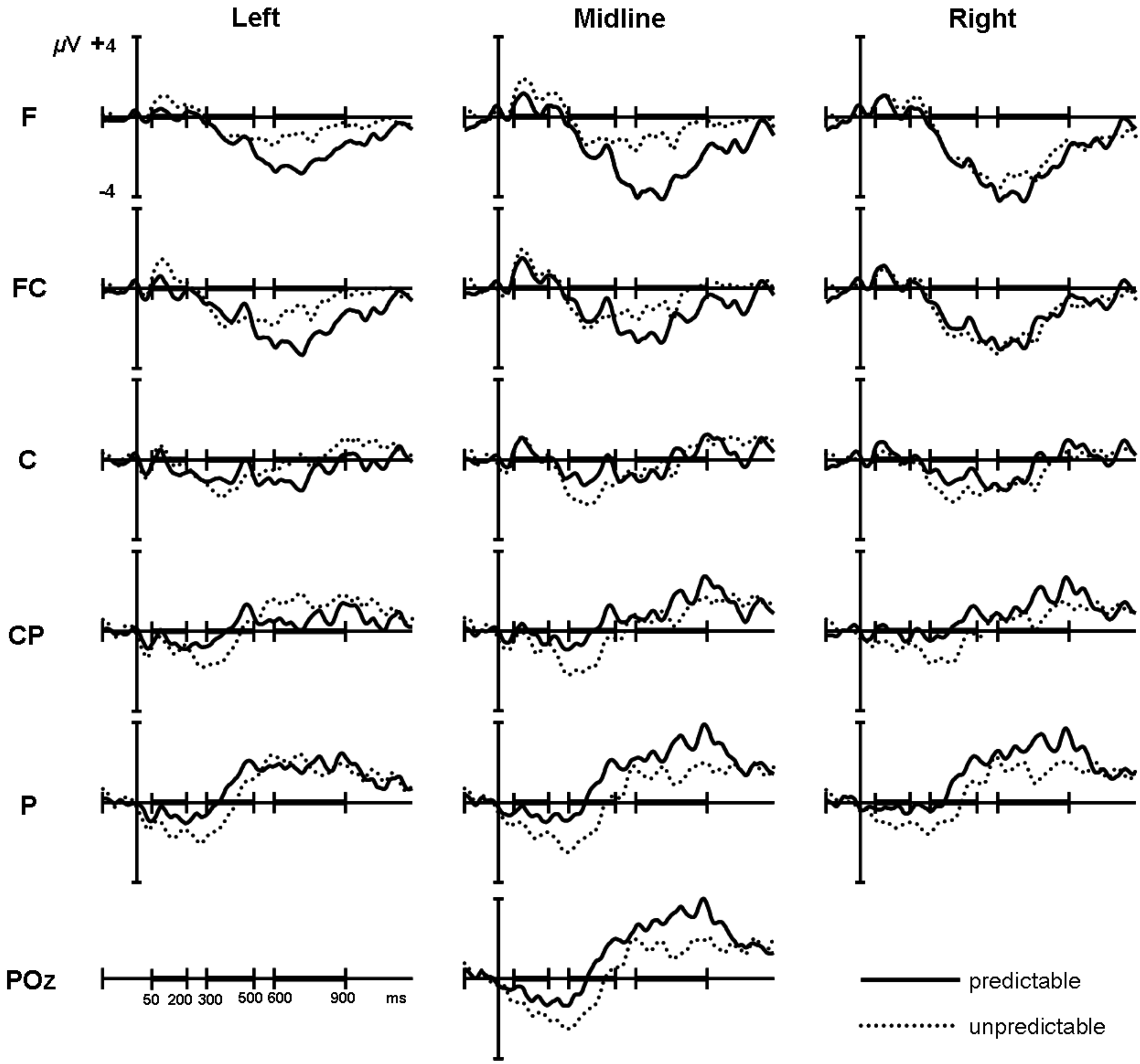
Wagner, A. D., Schacter, D. L., Rotte, M., Koutstaal, W., Maril, A., Dale, A. D., Rosen, B. R., Buckner, R. L., 1998. Building memories: Remembering and forgetting verbal experiences as predicted by brain activity. Science 281, 1188–1191.

Wilding, E. L., 2006. The practice of rescaling scalp-recorded event-related potentials. Biological Psychology 72, 325–332.

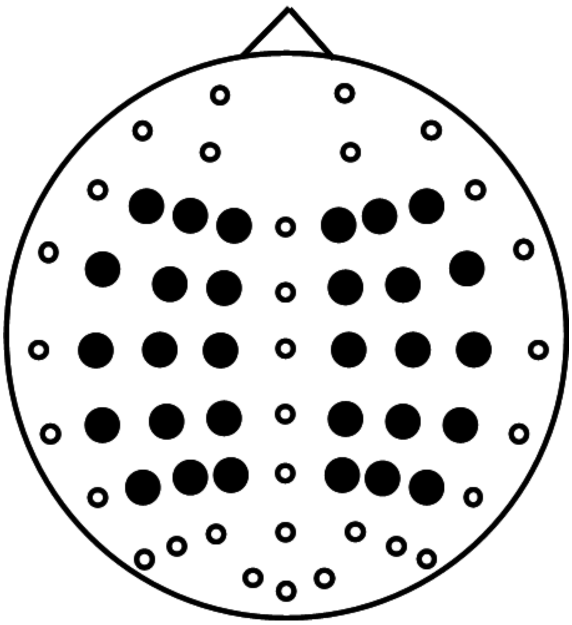
Zellner, B., 1994. Pauses and the temporal structure of speech. In: Keller, E. (Ed.), Fundamentals of speech synthesis and speech recognition. John Wiley, Chichester, pp. 31–62.

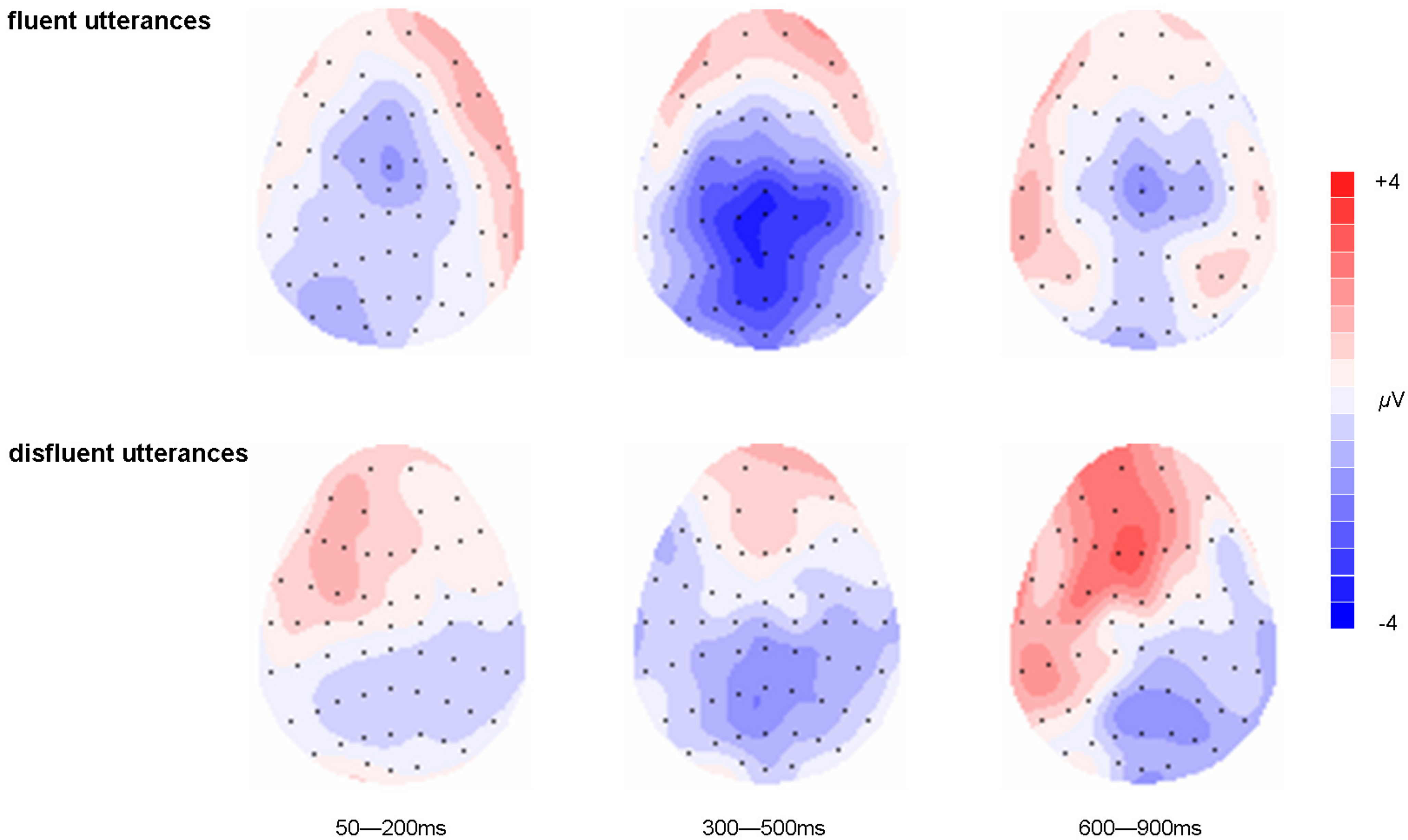


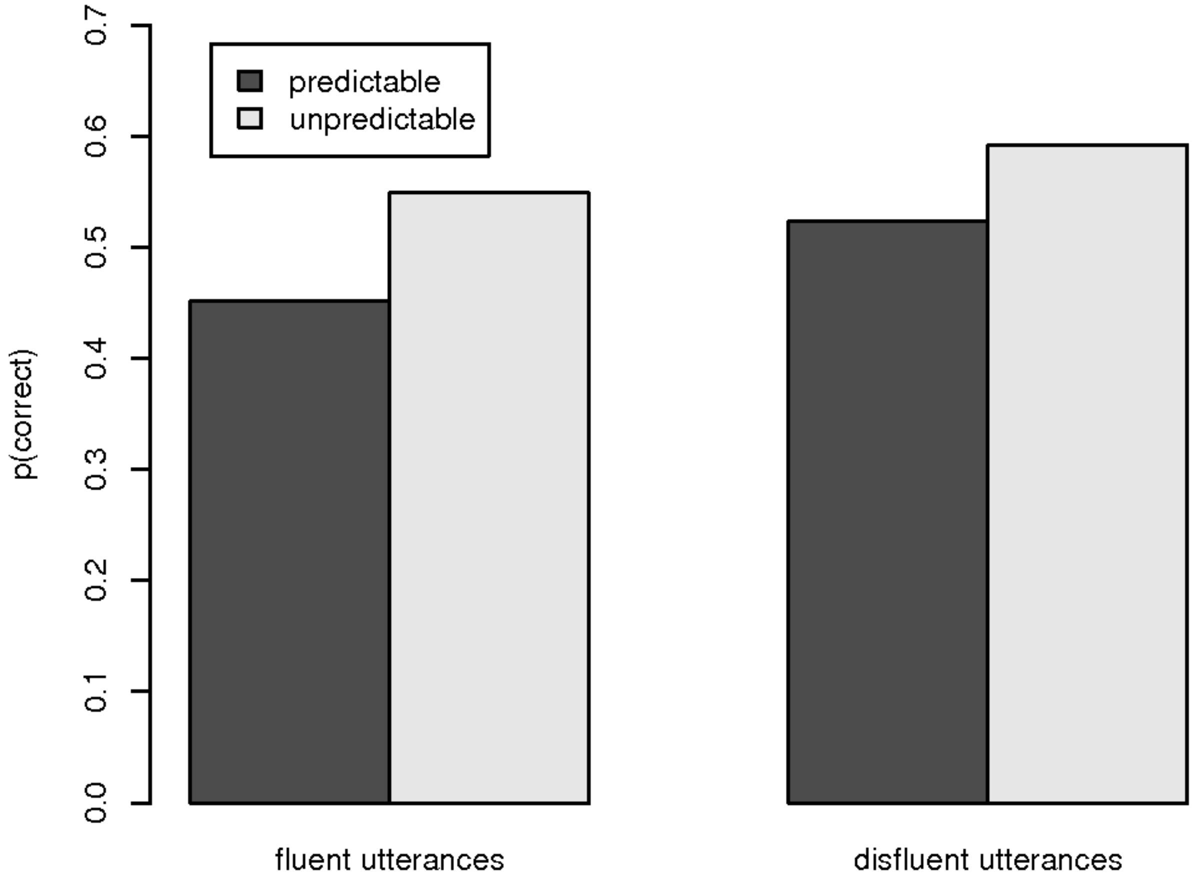
Left
Midline
Right
μV +4
-4
F
FC
C
CP
P
POz
50 200 300 500 600 900 ms
predictable
unpredictable



Left
Midline
Right
μV +4
-4
F
FC
C
CP
P
POz
50 200 300 500 600 900 ms
predictable
unpredictable




fluent utterances
disfluent utterances
50—200ms
300—500ms
600—900ms
+4
μV
-4



p(correct)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
predictable
unpredictable
fluent utterances
disfluent utterances