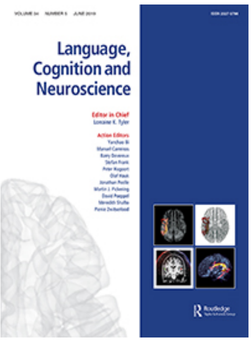

Routledge
Taylor & Francis Group

Language, Cognition and Neuroscience

ISSN: 2327-3798 (Print) 2327-3801 (Online) Journal homepage: https://www.tandfonline.com/loi/plcp21

Backward-looking sentence processing in typically disfluent versus stuttered speech: ERP evidence

Nathan D. Maxfield & Fernanda Ferreira

To cite this article: Nathan D. Maxfield & Fernanda Ferreira (2019) Backward-looking sentence processing in typically disfluent versus stuttered speech: ERP evidence, Language, Cognition and Neuroscience, 34:5, 561-579, DOI: 10.1080/23273798.2018.1533140

To link to this article: https://doi.org/10.1080/23273798.2018.1533140

Published online: 09 Oct 2018.

LANGUAGE, COGNITION AND NEUROSCIENCE
2019, VOL. 34, NO. 5, 561–579
https://doi.org/10.1080/23273798.2018.1533140

Routledge
Taylor & Francis Group

REGULAR ARTICLE

Check for updates

Backward-looking sentence processing in typically disfluent versus stuttered speech: ERP evidence

Nathan D. Maxfield[a] and Fernanda Ferreira[b]

[a]Department of Communication Sciences & Disorders, University of South Florida, Tampa, FL, USA; [b]Department of Psychology, University of California – Davis, Davis, CA, USA

ABSTRACT
The aim was to determine how backward-looking sentence processing is affected by typically disfluent versus stuttered speech. Two listener groups heard Garden Path (GP) and control sentences. GP sentences contained no disfluency, a silent pause, or a filled pause before the disambiguating verb. For one group, the sentence preambles additionally contained stuttering-like disfluencies. Comprehension accuracy, event-related potentials (ERPs) time-locked to disambiguating verbs, and perceptual speaker ratings, were compared between groups. The With Stuttering group perceived the speaker as less competent but had better comprehension accuracy for GP sentences. ERPs to disambiguating verbs in GP sentences included a P600 component, indexing backward-looking sentence processing, but only for the No Stuttering group. Other ERP components, elicited to GP sentences with silent/filled pauses, did not differ between groups. Results suggest that listeners abandon prior expectations when processing sentences containing stuttering-like disfluencies, possibly because they lack a speaker model defined by the presence of stuttering.

ARTICLE HISTORY
Received 15 April 2018
Accepted 28 September 2018

KEYWORDS
Sentence processing; stuttering; disfluencies; ERP; reanalysis

Introduction

Developmental stuttering, which persists in ~1% of adults, can impact quality-of-life (Yaruss, 2010) in domains of social, emotional and cognitive functioning (Beilby, Byrnes, Meagher & Yaruss, 2013; Craig, Blumgart, & Tran, 2009; Iverach et al., 2009). Speech therapy for adults who stutter (AWS) may aim to reduce struggle associated with stuttering and/or to reduce the frequency of stuttering (Blomgren, 2010; Prins & Ingham, 2009). However, typically there is little therapeutic focus on teaching AWS strategies for optimising comprehension of their listeners.

Existing evidence indicates that the frequency of stuttering can impact listener perceptions of and reactions toward stuttering, and that stuttering can interfere with listener recall and comprehension (Healey, 2010). Still needed is a comprehensive investigation of how stuttering impacts real-time sentence processing in oral comprehension. The aim of this study was to investigate how processing of ambiguous (Garden Path (GP)) sentences is affected by typically disfluent versus stuttered speech.

Listener perceptions and comprehension of stuttered speech: existing evidence

Wingate (1964) defined stuttering as primarily involving "audible or silent, repetitions or prolongations in the utterance of short speech segments, namely: sounds, syllables, and words of one syllable" (p. 488). Stuttering-like disfluencies may coincide systematically with certain word types (Brown, 1945; Howell, Au-Yeung, & Sackin, 1999; Kaasin & Bjerkan, 1982; Lanyon, 1969; Quarrington, 1956; Soderberg, 1971) and syntactic structures (Hannah & Gardner, 1968; Jayaram, 1984; Kaasin & Bjerkan, 1982; Koopmans, Slis, & Rietveld, 1991; Logan, 2003; Quarrington, 1956; Ratner & Benitez, 1985; Silverman & Ratner, 1997; Tornick & Bloodstein, 1976; Tsiamtsiouris & Cairns, 2013; Wells, 1979). An interesting question is how oral comprehension is affected by the presence of stuttering-like disfluencies.

In the earliest work of this type, sentences containing simulated single-unit syllable repetitions were presented to listeners. Sander (1965) found that sentence recall accuracy was reduced when listeners were told to focus on versus ignore the presence of stuttering. Hulit (1976) investigated the effects of severe (i.e. relatively frequent) repetition- and prolongation-type stuttering strategically located on key versus non-keywords. He found that the presence of both types of stuttering on non-keywords reduced sentence recall relative to fluent speech, while the presence of prolongation-type stuttering on keywords yielded sentence recall accuracy similar to that observed for fluent speech. Hulit suggested that the presence of prolongation-type stuttering on key-

CONTACT Nathan D. Maxfield nmaxfield@usf.edu

words had an effect of heightening attention toward rather than away from sentence information, in contrast to Sander's findings.

Cyprus, Hezel, Rossi, and Adams (1984) extended this line of work by comparing effects of mild (i.e. relatively infrequent) stuttering on sentence recall to severe stuttering in addition to a no-stuttering control condition. In this study, disfluencies were strategically placed on words of high or low information value (essentially, on content words or function words). College students recalled less material when severe stuttering took place on words of high information value – contrasting the results of Hulit (1976) – but recall in the mild stuttering versus no-stuttering conditions was similar. The authors concluded that severe stuttering can make the speech of an AWS difficult to remember, particularly on words with high semantic content.

More recently, Panico and Healey (2009) observed that free and cued recall of even mildly stuttered speech was impaired compared to fluent controls, and listeners indicated that more mental effort was required to comprehend stuttered speech. Earlier work from Healey's lab revealed that the presence of stuttering can impact listeners in other important ways too. Susca and Healey (2001) found that the more stuttering present in speech samples, the more negative the terms listeners used to describe the speech. In addition, listeners could apparently distinguish typically fluent from stuttered speech with all disfluencies and pauses removed, suggesting the presence of additional features that differentiate stuttered speech. Susca and Healey (2002) found that more severe stuttering was associated with perceptions that the speaker seemed flustered, less intelligent and less educated. Speech with more severe stuttering was also described as boring and more difficult to understand than speech with less stuttering. Interestingly, listeners noted that the speech was "hard to follow to predict next words", and that earlier portions of the speech were hard to remember (also see Panico, Healey, Brouwer, & Susca, 2005).

Disfluency effects on real-time sentence processing

A critical next step is to identify how stuttering-like disfluencies impact specific language and cognitive processes that support real-time sentence processing. The goal of spoken language comprehension is to build a coherent mental representation based on input provided by the speaker. We assume that sentence comprehension is an active process by which listeners construct an initial interpretation based on prediction and world knowledge, and continuously improve and modify it by incorporating new information moment-by-moment until a likely or useful interpretation is reached.

In a recent study (Lowder, Maxfield & Ferreira, submitted), we used a visual-world eye-tracking paradigm to investigate how predictive sentence processing is affected when the speaker also produces atypical disfluencies (i.e. stuttering). Participants heard sentences containing self-repair disfluencies and control sentences while viewing visual displays containing a predictable target object that was never actually named. Half the participants heard the sentences spoken by a speaker who stuttered mildly (once or twice) in the preamble of each sentence, while the other half of the participants heard the same sentences spoken by the same speaker using controlled fluency (learned in speech therapy). Results indicated that listeners' ability to model the production system of the speaker when he stuttered was disrupted, as was listeners' ability to engage in predictive language processing.

Crucially, when predictions are incorrect in sentence processing, ambiguities may arise that need to be resolved via backward-looking sentence processing (e.g. reanalysis/updating of sentence interpretations). GP sentences provide a context for investigating this type of processing. In a sentence such as, "While the man hunted the deer ran into the woods" (Bailey & Ferreira, 2003), a listener may initially treat the second noun phrase, *the deer*, as the object/theme of the subordinate clause verb, *hunt*. However, upon encountering the second verb, *deer*, the listener must revise the initial interpretation, as *the deer* now serves as the obligatory subject/agent of the matrix clause. Bailey and Ferreira (2003) found that listeners were more likely to judge sentences as ungrammatical when a typical disfluency (silent or filled pause) was located just before the second (disambiguating) verb in GP sentences. In these conditions, listeners seemed to commit to the initial, erroneous sentence interpretation rather than engaging backward-looking sentence processing.

In a study using brain event-related potentials (ERPs), Maxfield, Lyon, and Silliman (2009) found that the P600 – an ERP index of backward-looking sentence processing (see Frisch, Schlesewsky, Saddy, & Alpermann, 2002) – was undetectable to disambiguating verbs in GP sentences when those verbs were preceded by typical disfluencies (e.g. *While the man hunted the deer uh uh <u>jumped</u> over the fence*). Typically, such verbs would elicit P600, marking the parser's attempt to revise syntactic ambiguity. Absence of P600 activation suggests that the presence of typical disfluencies in GP sentences forestalled backward-looking sentence processing when the disambiguating verb was presented, possibly because the disfluencies forced listeners to "linger" on the initial (erroneous) parse for so long that they committed to it. On the other hand, those same verbs did elicit the N400 – an ERP index of semantic integration (Kutas & Federmeier, 2011). N400 activation suggests that listeners not only accepted their original (erroneous) parses for material preceding the disambiguating verbs, but tried integrating the disambiguating verbs into those parses. Disambiguating verbs preceded by typical disfluencies also elicited a left anterior ERP component, possibly marking that the disfluencies cued listeners to heighten attention toward the disambiguating verbs and/or to give them special status (see Maxfield et al., 2009).

The current study aims to determine whether backward-looking sentence processing is affected in the same way when speech contains stuttering-like disfluencies in addition to typical disfluencies. Two different models of sentence processing provide differing predictions. According to the Ambiguity Resolution model, verbal input is broken down linguistically to arrive at an interpretation and this process is not speaker-dependent (see Hanulíková, Van Alphen, Van Goch, & Weber, 2012). From this perspective, the effect outlined above – P600 attenuation to disambiguating verbs in GP sentences when those verbs are preceded by typical disfluencies – should be observed even if a speaker is identified as someone who stutters. Alternatively, the Noisy Channel model of sentence processing suggests that comprehension includes a mechanism for normalising improbable input caused by "imperfections" in signal transmission (Gibson, Bergen, & Piantadosi, 2013), including speaker error. From this perspective, typical disfluencies in GP sentences may not have a disrupting effect on backward-looking sentence processing if a speaker is identified as someone who stutters. Instead, typical disfluencies may be given a different status. For example, silent and filled pauses – which often appear in the speech of people who stutter as strategic devices for managing stuttering – may be perceived as "place holders" signalling continuations in discourse. In turn, listeners may continue actively parsing sentences beyond those disfluencies. If so, disambiguating verbs preceded by typical disfluencies in GP sentences might still trigger backward-looking sentence processing, in which case P600 activation might still be observed to those verbs.

Methods

Participants

Two groups of 15 undergraduate students participated in the study ($n = 8$ females per group). All 30 participants were recruited from the University of South Florida Psychology Department Participant Pool and provided written informed consent before participating. All study procedures were approved and monitored by the University of South Florida Institutional Review Board. One group listened to stimuli containing no stuttering (mean age = 19 years, 8 months), while the second group listened to stimuli containing stuttering (mean age = 19 years, 9 months). All participants were monolingual speakers of Standard American English. They and their parents were born in the United States. All participants reported that neither they nor anyone in their immediate family suffered from speech, language, hearing or learning difficulties, including stuttering. All self-reported right-hand dominance and indicated right-hand preference on the Edinburgh Handedness Inventory (Oldfield, 1971). On the date of testing, each participant reported being in good health and not taking medication that can affect cognitive functioning. No participant reported a history of neurological injury or disease. All participants had normal or corrected-to-normal vision, and all passed a pure-tone audiometric screening on the date of testing. To pass, participants were required to respond reliably to 20 dBHL tones presented by a GSI-17 audiometer at 1000, 2000, and 4000 Hz, respectively. Finally, none of the participants had taken coursework in linguistics or communication sciences and disorders.

Materials

Ninety GP sentences and 90 non-GP sentences were presented to each listener Group (see Table 1 for a summary). Full details about the sentence designs, and

Table 1. A summary of the six sentence conditions presented to each listener group.

Sentences Presented to the No Stuttering Listener Group ($n = 180$)
- GP sentences ($n = 30$)
- GP sentences with a Silent Pause before the disambiguating verb ($n = 30$)
- GP sentences with a Filled Pause before the disambiguating verb ($n = 30$)
- non-GP sentences containing no Silent or Filled Pauses (Control) ($n = 30$)
- non-GP sentences with a Silent Pause between the second NP and the matrix clause verb ($n = 30$)*
- non-GP sentences with a Filled Pause between the second NP and the matrix clause verb ($n = 30$)*

Sentences Presented to the With Stuttering Listener Group ($n = 180$)
- GP sentences with stuttering in the preamble ($n = 30$)
- GP sentences with stuttering in the preamble and a Silent Pause before the disambiguating verb ($n = 30$)
- GP sentences with stuttering in the preamble and a Filled Pause before the disambiguating verb ($n = 30$)
- non-GP sentences with stuttering in the preamble but no Silent or Filled Pauses (Control) ($n = 30$)
- non-GP sentences with stuttering in the preamble and a Silent Pause between the second NP and the matrix clause verb ($n = 30$)*
- non-GP sentences with stuttering in the preamble and a Filled Pause between the second NP and the matrix clause verb ($n = 30$)*

Note: Asterisks indicate sentence conditions that were included in the experiment but not included in analyses of probe question accuracy or ERP effects.

the sentences themselves, are provided in Maxfield et al. (2009, see Methods and Appendix A). The 90 GP sentences included: (a) 30 GP sentences with no filled or unfilled pauses directly preceding the disambiguating verb; (b) 30 GP sentences with an unfilled (silent) pause directly before the disambiguating verb; and (c) 30 GP sentences with a filled pause ("uh uh") directly preceding the disambiguating verb.

The 90 non-GP sentences were created by replacing the transitively biased verb in the subordinate clause with an intransitive verb in each of the 90 GP sentences. Thirty non-GP sentences did not contain a silent or filled pause and served as Control items in this experiment. Thirty non-GP sentences contained a silent pause, and another 30 contained a filled pause, between the second NP and the matrix clause verb. These latter 60 items were included to ensure that the probability of hearing sentences containing silent or filled pauses was equal across GP and non-GP conditions (i.e. we did not want listeners to associate the presence of silent or filled pauses with GP sentences only). However, probe question accuracy and ERP data from the 60 non-GP sentences containing silent or filled pauses were not ultimately analysed.

Sentence recording

A 23-year-old Caucasian male, monolingual speaker of Standard American English with a clinical diagnosis of stuttering since childhood, was recruited to generate the verbal stimuli. Following years of speech therapy, the speaker was able to converse and read aloud with controlled fluency. Fluency controls were subtle and primarily involved (a) initiating the first word of each speech group with slightly exaggerated articulatory movements and gentle voicing, and (b) maintaining fluency in each speech group by sustaining air flow and using soft articulatory contacts. The speaker was also able to speak with fluency controls off, during which stuttering was mild to moderate in severity (i.e. somewhat frequent with some stuttering-associated struggle behaviour).

The speaker was recorded at a sampling rate of 44.1 kHz using a Sony digital audio tape recorder and high-quality microphone inside a sound-attenuating booth. To control for disambiguating prosodic cues in the GP conditions, he read all GP sentences with a carrier phrase "According to Mary … " (see Bailey & Ferreira, 2003). This allowed the reader to maintain constant prosody while reading each GP sentence instead of pausing (with sharply rising and then falling intonation) before the object noun phrase. For each sentence containing a typical disfluency, the speaker added an unfilled pause (silently counting to two) or a filled pause ("uh uh") before the critical verb.

In the No Stuttering condition, the speaker produced the sentences with controlled fluency. These sentences were closely scrutinised by the first author, a Speech-Language Pathologist with extensive experience in assessment of fluency disorders, to verify that stuttering-like disfluencies were not present. In the With Stuttering condition, the speaker produced each sentence with stuttering-like disfluencies in the sentence preamble. These disfluencies consisted of sound repetitions or sound prolongations (silent blocks, a third type of stuttering-like disfluency, were never produced by the speaker). Each stuttered sentence always contained one, but no more than two, instances of stuttering-like disfluency in the sentence preamble (i.e. before the critical verb).[1] Crucially, stuttering-like disfluencies were never time-locked with the critical verb in each sentence or with the word directly preceding it. Furthermore, there were never instances of stuttering following the critical verb.

Analysis of stuttering frequency for sentences in the With Stuttering condition revealed that non-GP (Control) sentences contained 8.3% stuttered syllables; GP sentences without a typical disfluency preceding the critical verb contained 9.05% stuttered syllables; GP sentences with an unfilled pause preceding the critical verb contained 8.2% stuttered syllables; and GP sentences with a filled pause preceding the critical verb contained 8.4% stuttered syllables. Percentage of syllables stuttered was computed for filled pause conditions without including "uh uh" in the frequency counts.

The sentences were edited offline using Sony SoundForge software. First, the continuous digital audio recording was imported to SoundForge, maintaining the 44.1 kHz sampling rate to ensure high fidelity in the recordings. Next, each sentence was segmented (i.e. "spliced") from the continuous recording with no silence preceding the first word or following the last word. Third, the stimuli were normalised so that the peak root mean square (RMS) amplitude (in dB) was the same for each sentence. Finally, the duration (in milliseconds) from the onset of each sentence to the onset of the critical verb, the total duration of each sentence, and the difference between these two points, was measured.

The average duration from the onset of the critical verbs to the end of the sentences was compared for eight conditions (No Stuttering Control sentences, 914 ms (SD = 171); With Stuttering Control sentences, 887 ms (SD = 170); No Stuttering GP sentences with no filled or unfilled pause preceding the critical verb, 973 ms (SD = 209); With Stuttering GP sentences with no filled or unfilled pause preceding the critical verb, 899 ms (SD = 194); No Stuttering GP sentences with unfilled pause preceding the critical verb, 920 ms (SD = 223); With Stuttering GP sentences with unfilled pause preceding the critical verb, 922 ms (SD = 220); No Stuttering GP sentences with filled pause preceding the critical verb, 921 ms (SD = 198); and With Stuttering GP sentences with filled pause preceding the critical verb, 920 ms (SD = 196)). A main effect of Condition was not detected ($F[7,239] = 0.482$, $p = .85$), indicating that the durations of the critical verbs plus the small number of words following them was relatively well-matched between conditions. The maximum duration from the onset of the critical verb to the end of the sentence in any condition was 1469 ms.

Probe questions

One Yes/No probe question was designed for each sentence. Each question tested whether the second NP was interpreted as the object/theme of the subordinate clause, or as the subject/agent of the matrix clause (following Christianson, Hollingworth, Halliwell, & Ferreira, 2001; Maxfield et al., 2009). Questions regarding the matrix clause were included to determine whether subjects completed some level of reanalysis, enough to answer questions regarding the role of the second NP as subject/agent of the matrix clause correctly. Questions about the role of the NP in the subordinate clause were included to determine whether subjects dropped their original interpretation of the NP as object/theme where appropriate (i.e. in the GP conditions). In each condition, half the questions probed the subordinate clause and half probed the matrix clause.[2]

Procedure

The main testing procedure was identical to that used in Maxfield et al. (2009). Participants were instructed to listen carefully to each sentence, answer a Yes/No question by pressing one of two buttons on a response box, and rate their confidence along a 4-point scale. Each session lasted ~30 minutes and included 5 blocks of 36 sentences. Each block contained six sentences from each of the six different conditions, presented in randomised order. A crosshair (+) was displayed on the computer monitor as each sentence was presented. A written probe question appeared 2000 ms after the offset of the sentence-final word and remained on the screen until the participant responded. Next, confidence was rated, followed by an intertrial interval of 1500 ms.

Immediately following the task, participants evaluated the speaker. Perceptual evaluations were collected primarily to determine whether the speaker was perceived differently in the With Stuttering versus No Stuttering conditions. We also explored whether perceptual evaluations of the speaker were associated with probe response accuracy in either speaker condition. First, each participant was asked, "How would you describe this person's speech?", and responses were recorded in writing. In addition, participants responded using a 7-point Likert scale (strongly disagree to strongly agree) to each of four statements: (1) This person is a competent speaker, (2) This person is a fluent speaker, (3) This person read the sentences easily, and (4) I felt comfortable listening to this person. Finally, participants provided open-ended responses to the following questions: "What contributed to how comfortable you felt listening to this person?"; "How easy or difficult was it to understand the sentences you just heard?"; and "Did the way this person spoke interfere with your understanding of the sentences?". Responses to these last few questions were quite diverse and will not be reported here.

Apparatus and recording

Participants were tested in a dimly lit, sound-attenuating booth facing a 19-inch computer monitor. They heard the sentence stimuli through insert earphones (Etymotic Research, Model E-2). Probe responses and confidence ratings were registered using a push-button response box (Psychological Software Tools). E-prime experimental control software (Psychological Software Tools, version 1.1) was used to present the sentences and log behavioural responses.

Continuous EEG was recorded from each participant at a sampling rate of 500 Hz using SCAN software, Version 4.3 (Neuroscan). Each participant wore a nylon QuikCap (Neuroscan) fitted with 62 active recording electrodes positioned following the International 10–20 system (Klem, Lüders, Jasper, & Elger, 1999). Active recording electrodes were referenced to a midline vertex electrode. A ground electrode was positioned on the midline, anterior to Fz. Two bipolar-referenced vertical electro-oculograph (VEOG) electrodes, and two bipolar-referenced horizontal electro-oculograph (HEOG) electrodes, recorded electro-ocular activity. Electrodes were constructed of Ag/AgCl. Electrode impedance was kept below 5 kOhms. Continuous EEG was low-pass filtered online at a corner frequency of 100 Hz (time constant: DC). E-prime software sent a trigger to the EEG file at the onset of each sentence and also at the onset of the critical (second) verb in each sentence.

EEG-to-average-ERP data reduction

Replicating the EEG data processing sequence in Maxfield et al. (2009), first the continuous EEG record of each participant was epoched. Each epoch contained EEG data recorded from each of 62 active recording electrodes, time-locked to the critical (second) verb in each sentence, starting 200 ms before verb onset and terminating 2200 ms after verb onset. The epoch duration was later truncated to a target time interval (0–2000 ms relative to verb onset) following averaging. However, an extended epoch duration was used at first to ensure that the procedures, described next, would adequately correct or reject artefacts on the leading and trailing edges of the target (0–2000 ms) time interval.

EEG ocular artefact correction

In order to include as many trials as possible in EEG averaging (Picton et al., 2000), we used an Independent Component Analysis (ICA)-based (Bell & Sejnowski, 1995) ocular artefact correction procedure modified from Dien (2010). At least one blink component was identified for each participant. The approach used here has been shown to accurately identify and remove ocular artefact without significantly warping/skewing ERP activity (Glass et al., 2004).

EEG trial rejection

After ICA blink correction, waveforms at each active recording electrode were checked for noise, separately for each trial. Any electrode for which the fast-average amplitude exceeded 200 microvolts (large drift) was marked bad, as was any electrode at which the differential amplitude exceeded 100 microvolts (high-frequency noise). Any trial with more than three bad electrodes was rejected. No participant lost more than ∼17% of trials for any condition due to bad channel artefact, and most lost well under 10% of trials per condition.

Final EEG processing

For any accepted trial with electrodes marked bad (≤3), the EEG activity at those electrodes was replaced using spherical spline interpolation (Nunez & Srinivasan, 2006, Appendices J1–J3). The EEG trials were then averaged together, separately for each condition. As a result, each participant had four sets of ERP averages, one for each condition targeted for analysis. For each participant, no fewer than 25 artefact-free trials went into the set of ERP averages for each condition. The averaged ERP data were truncated to the target time interval (0–2000 ms after verb onset), re-referenced to linked mastoids, and baseline-corrected using a post-stimulus baseline (0 to +100 ms relative to verb onset).[3] Typically, P600 onset is at ∼500 ms and its peak latency at ∼800 ms. However, P600 onset and peak latency can vary with the time needed for comprehenders to diagnose and revise an erroneous parse (Friederici, 1998). To make sure any P600 effects were detected, the target time interval (0–2000 ms) spanned the entire duration between verb onset and probe question onset.

Data analysis

Listener perception data

Responses to the question, "How would you describe this person's speech?", were compared and contrasted between Groups. Ratings to each of the four Likert-scale questions were compared between Groups using a Mann–Whitney *U* test. *U* tests were two-sided with an alpha level of 0.05.

Behavioural data

Probe question accuracy was scored automatically during testing for each participant by E-prime software. Accuracy rates were submitted to a repeated-measures analysis of variance (ANOVA). Sentence Type was treated as a within-subjects factor with four levels (Control, GP, GP + silent pause, GP + filled pause), Clause Type was treated as a within-subjects factor with two levels (matrix versus subordinate), and Group was treated as a between-subjects factor with two levels (No Stuttering versus With Stuttering). The ANOVA was two-sided with an alpha level of 0.05. A main effect of Condition or interaction involving Condition was followed with *t*-tests comparing each GP condition with Control. The alpha level for declaring statistical significance was set at $p < .0167$ (0.05/3 comparisons versus Control).

Correlations between listener perceptions and probe response accuracy

Spearman rho correlations were computed between (a) responses to each Likert-scale question and (b) probe response accuracy to each Clause Type in each Sentence Type, separately for each Group.

Electrophysiological data

Following the procedures used in Maxfield et al. (2009), the set of averaged ERP data was submitted to a covariance-based, two-step, temporal-spatial PCA (Dien & Frishkoff, 2005). In the first step, the ERP averages were entered into a matrix with 1001 columns (one column per sampling point between 0 and 2000 ms including 0 ms, with sampling points occurring once every 2 ms) and 7440 rows (averaged ERPs for each of 15 participants in each of two groups, at each of 62 electrodes, in each of 4 conditions). This matrix was submitted as input to a temporal PCA to identify distinct windows of time in the ERP averages (temporal factors) during which similar voltage variance was active across consecutive sampling points. Twenty-five temporal factors were retained. In the second step, a spatial PCA was performed on the factor scores associated with selected temporal factors. The scores for each temporal factor (representing the ERP variance within a specific time window) were entered into a matrix with 62 columns (one column per electrode) and 120 rows (temporal factor scores for 30 participants, in each of 4 conditions). Each of these matrices was submitted as input to a spatial PCA to identify topographically coherent regions of ERP variance (spatial factors) within the time window associated with each temporal factor. Four spatial PCAs were carried out, one for each of the four targeted temporal factors.

The following specific procedures were used to conduct the initial temporal PCA, and each of the subsequent spatial PCAs. Rule M (Preisendorfer & Mobley, 1988) was used to determine how many dominant-variance components could be extracted from each matrix. Components meeting this criterion were rotated to simple structure using Promax (Hendrickson & White, 1964) with Kaiser normalisation and $k = 2$ (Richman, 1986; Tataryn, Wood, & Gorsuch, 1999). All PC analyses were completed using PCA Toolbox (Dien, 2010). These procedures replicate those used in Maxfield et al. (2009) to analyse ERP effects. As with that analysis, one goal of the current analysis was to identify a temporal-spatial factor combination consistent with P600 activation (based on its time-course, scalp topography, and associated variance). However, since other ERP effects of interest were detected in Maxfield et al. (2009), the current data set was explored for effects beyond P600 too. To test for experimental effects, factor scores summarising the voltage variance associated with specific pairs of temporal and spatial factors were submitted to repeated-measures ANOVA with Sentence Type as a within-subjects factor with four levels (Control, GP, GP + silent pause, GP + filled pause) and Group as a between-subjects factor with two levels (No Stuttering versus With Stuttering). Temporal-spatial factor combinations that were interpretable both in terms of time-course and scalp topography, and explained at least 1% of the variance, were targeted. ANOVAs were two-sided with an alpha level of 0.05. Reported *p*-values were corrected when the assumption of sphericity was violated (Greenhouse & Geisser, 1959). A main effect of Condition or interaction involving Condition was followed with *t*-tests comparing each GP condition with Control. The alpha level for declaring statistical significance of post-hoc *t*-tests was $p < .0167$ (0.05/3 comparisons versus Control).

Results

Listener perception data

When asked, "How would you describe this person's speech?", 14 listeners in the With Stuttering group used the word "stutter" and one listener used the word "broken" to describe the speech. Five listeners in the No Stuttering group used the word "stutter" to describe the speech while the other 10 listeners focused on the presence and quality of pauses in the sentences.

Numerical ratings to each of the four Likert-scale questions differed between Groups. Listeners in the No Stuttering group were more likely than listeners in the With Stuttering group to agree that: (a) the speaker was competent (mean = 4.4, SD = 1.55 versus mean = 2.73, SD = 1.39) ($U = 48.5$, $p = .007$); (b) the speaker was fluent (mean = 4.33, SD = 1.5 versus mean = 3.1, SD = 1.79) ($U = 65$, $p = .043$); (c) the speaker read the sentences easily (mean = 4.13, SD = 1.25 versus mean = 2.1, SD = 0.8) ($U = 20.5$, $p < .001$); and (d) they felt comfortable listening to the speaker (mean = 4.6, SD = 1.9 versus mean = 2.93, SD = 1.28) ($U = 53.5$, $p = .01$).

Behavioural data

As shown in Figure 1, probe questions to Control sentences were answered more accurately than to other sentence types in both Groups and to both Clause Types. Repeated-measures ANOVA revealed a main effect of Clause Type ($F[1,28] = 38.93$, $p < .001$), a main effect of Sentence Type ($F[3,84] = 57.43$, $p < .001$), and an interaction of Sentence and Clause Type ($F[3,84] = 22.82$, $p < .001$). For Subordinate clause questions, responses to Control sentences were more accurate than to GP ($p < .001$), GP + silent pause ($p < .001$) and GP + filled pause ($p < .001$) sentences. Additionally, responses to GP sentences were more accurate than to GP + silent pause ($p = .003$) and GP + filled pause ($p = .002$) sentences. For Matrix clause questions, responses to Control sentences were also more accurate than to GP ($p = .001$), GP + silent pause ($p < .001$) and GP + filled pause ($p < .001$) sentences. Additionally, responses to GP sentences were more accurate than to GP + silent pause ($p < .001$) and GP + filled pause ($p < .001$) sentences. Responses did not differ between GP + silent pause and GP + filled pause for either Subordinate ($p = .35$) or Matrix ($p = .09$) clause questions. Crucially, between clause types, answers were more accurate to Matrix versus Subordinate clause questions in GP ($p < .001$), GP + silent pause ($p = .001$) and GP + filled pause ($p < .001$) sentences.

Probe question accuracy was also affected by a main effect of Group ($F[1,28] = 7.19$, $p = .01$) and an interaction of Group and Sentence Type ($F[3,84] = 5.35$, $p = .02$). Both Groups had more accurate responses to Control sentences than to GP, GP + silent pause, and GP + filled pause sentences ($p < .001$ for each pairwise test versus Control in each Group). Crucially, between groups, sentences With Stuttering were answered more accurately than No

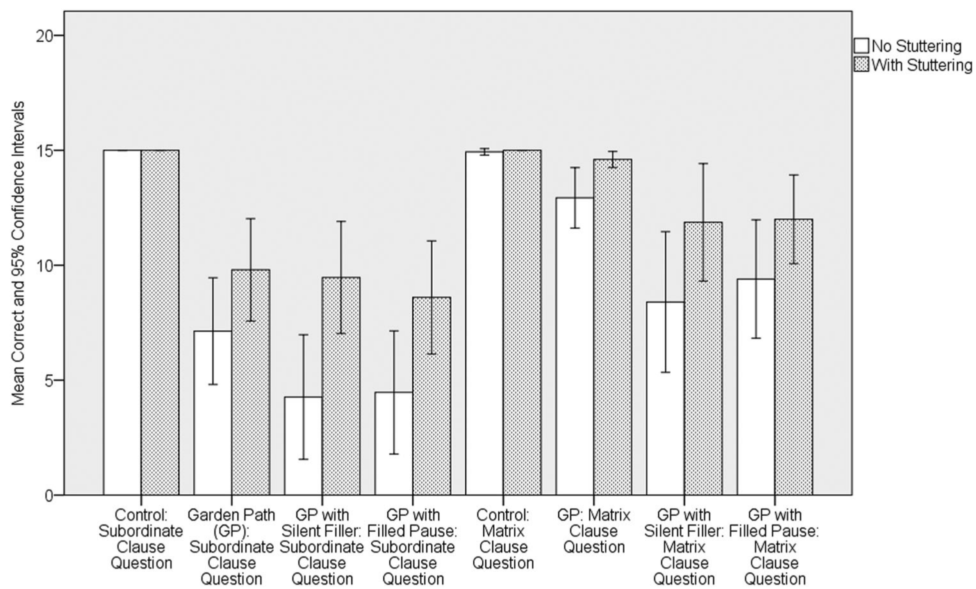

Figure 1. Mean number of items correct and 95% confidence intervals for each Group in each Condition.

Stuttering sentences in the GP ($p = .02$), GP + silent pause ($p = .01$), and GP + filled pause ($p = .03$) conditions.

Correlations between listener perceptions and probe response accuracy

Statistically significant, positive Spearman rho correlations were found between Likert-scale ratings of the speaker and probe response accuracy for the No Stuttering group only. Specifically, responses to the first Likert-scale question ("This person is a competent speaker") were positively correlated with response accuracy to probe questions about the Matrix clause in GP + silent pause sentences ($r_s = .59$, $p = .02$). Responses to the second Likert-scale question ("This person is a fluent speaker") were positively correlated with response accuracy to probe questions about the Subordinate clause in GP sentences ($r_s = .53$, $p = .04$), and also with response accuracy to probe questions about the Matrix clause in GP ($r_s = .7$, $p = .004$), GP + silent pause ($r_s = .71$, $p = .003$) and GP + filled pause ($r_s = .71$, $p = .003$) sentences. Finally, responses to the third Likert-scale question ("This person read the sentences easily") were positively correlated with response accuracy to probe questions about the Matrix clause in GP + silent pause ($r_s = .61$, $p = .02$) and GP + filled pause ($r_s = .56$, $p = .03$) sentences.

Brain electrophysiological data

Grand average waveforms to each condition are shown at each of 15 electrodes for the No Stuttering group in Figure 2, and for the With Stuttering group in Figure 3.

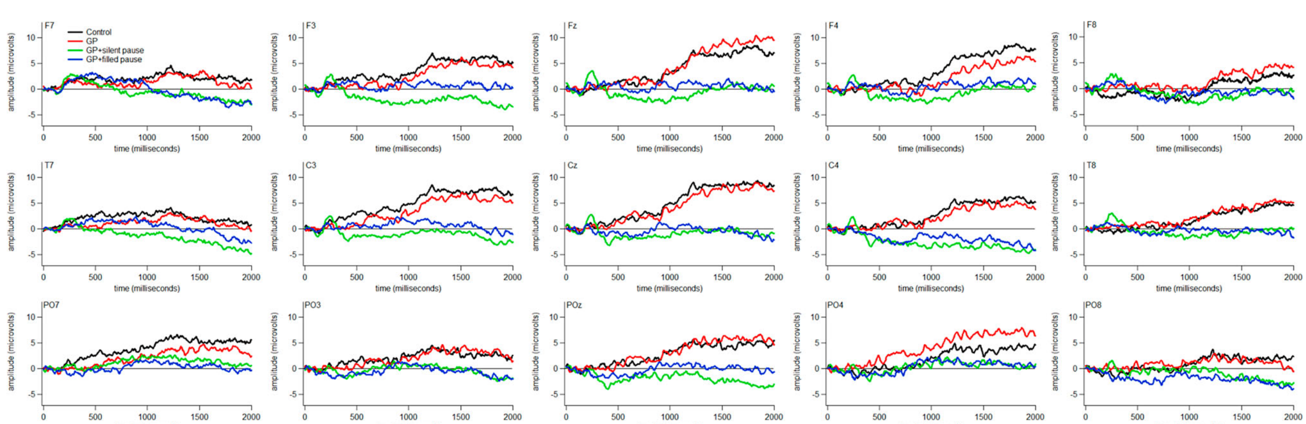

Figure 2. Grand average waveforms for the No Stuttering group, to each of the 4 main conditions (black = Control, red = GP, green = GP + silent pause, blue = GP + filled pause), at each of 15 electrodes (left to right, first row electrodes: F7, F3, Fz, F4, F8; second row electrodes: T7, C3, Cz, C4, T8; third row electrodes: PO7, PO3, POz, PO4, PO8). See Figures 4 through 7 for zoomed-in plots at specific electrodes.

Note that these 15 electrodes are shown for display purposes, and the entire set of electrodes was included in the analysis reported below. For both Groups, visual inspection suggested a number of differences in ERP activity elicited to the GP versus Control conditions. Most notable is a putative P600 effect to GP versus Control in the No Stuttering group (see Figure 2, electrodes POz, P04, P08 at ~700 ms after verb onset) which seemed attenuated or absent to the GP + disfluency conditions in this same group. This effect also seemed attenuated or absent to all GP conditions in the With Stuttering group (see Figure 3, posterior electrodes).

The ERP data set was mined using a two-step, sequential temporal-spatial PCA. For the initial temporal PCA, 25 temporal factors were Promax-rotated, accounting for 79.91% variance. Next, spatial PCA was carried out on the factor scores associated with specific temporal factors to identify scalp regions of coherent ERP activity within the time window defined by each temporal factor. Temporal-spatial factor combinations associated with statistically significant effects (after alpha-correction to control for Type-1 error, as described previously) are reported next. Statistically significant effects were detected in the virtual time windows associated with four temporal factors. Factor loadings, depicting the time-course of each temporal factor, are illustrated at the top of Figures 4 through 7. Each temporal factor will, hereafter, be labelled according to its peak latency, defined by the highest loading of each temporal factor (i.e. T254, T460, T648, and T1196, respectively).

T254, right anterior-temporal activity

Three spatial factors were Promax-rotated for T254, partitioning the voltage variance within the T254 time window into three scalp regions. One resulting spatial factor had a right anterior-temporal scalp topography (see Figure 4). Temporal-spatial factor scores – summarising ERP activity at ~254 ms after verb onset at the right anterior-temporal scalp region – were affected by Sentence Type ($F[3,84] = 5.99$, $p = .001$). Bonferroni-corrected *t*-tests revealed that activity to GP + silent pause was positive-going versus Control ($p = .002$). This same effect was detected for GP + filled pause versus Control ($p = .01$) (see Figure 4). Positive-going activity to the two GP + disfluency conditions is visually evident at ~250 ms in the grand averages of both groups (e.g. at electrode F8 in Figures 2 and 3).

T460, posterior activity

Three spatial factors were Promax-rotated for T460, partitioning the voltage variance within this time window into three scalp regions. One spatial factor had a posterior scalp topography (see Figure 5). Temporal-spatial factor scores – summarising ERP activity at ~460 ms after verb onset at the posterior scalp region – were affected by Sentence Type ($F[3,84] = 8.95$, $p < .001$). Bonferroni-corrected *t*-tests revealed that activity to GP + silent pause was negative-going versus Control ($p < .001$) as was activity to GP + filled pause ($p = .001$) (see Figure 5). Negative-going activity to the two GP + disfluency conditions is visually evident at ~460 ms in the grand averages of both groups (e.g. at electrode POz in Figures 2 and 3).

T648, posterior activity

Three spatial factors were Promax-rotated for T648, partitioning the voltage variance in this time window into three scalp regions. One spatial factor was defined by a posterior scalp topography (see Figure 6). Temporal-

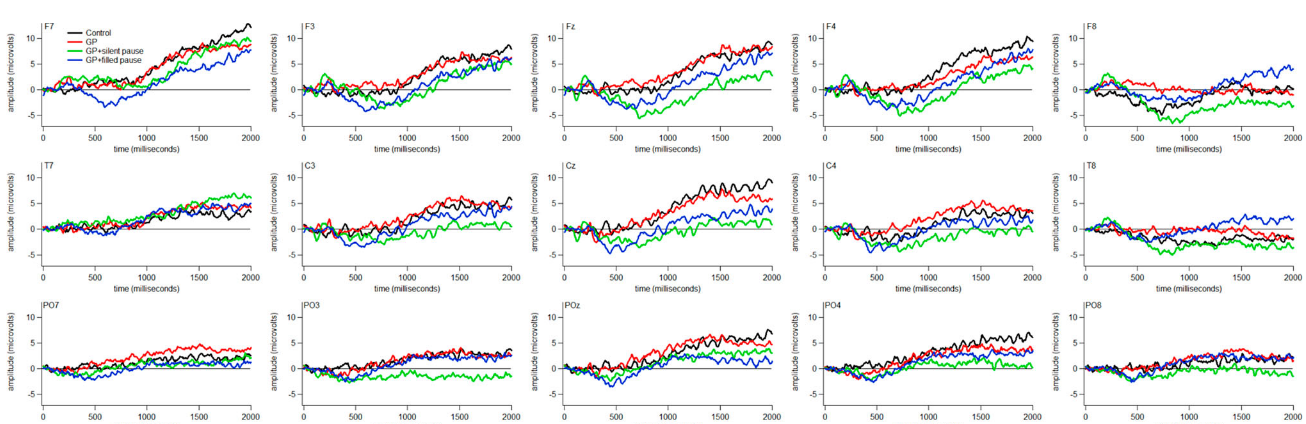

Figure 3. Grand average waveforms for the With Stuttering group, to each of the 4 main conditions (black = Control, red = GP, green = GP + silent pause, blue = GP + filled pause), at each of 15 electrodes (left to right, first row electrodes: F7, F3, Fz, F4, F8; second row electrodes: T7, C3, Cz, C4, T8; third row electrodes: PO7, PO3, POz, PO4, PO8). See Figures 4 through 7 for zoomed-in plots at specific electrodes.

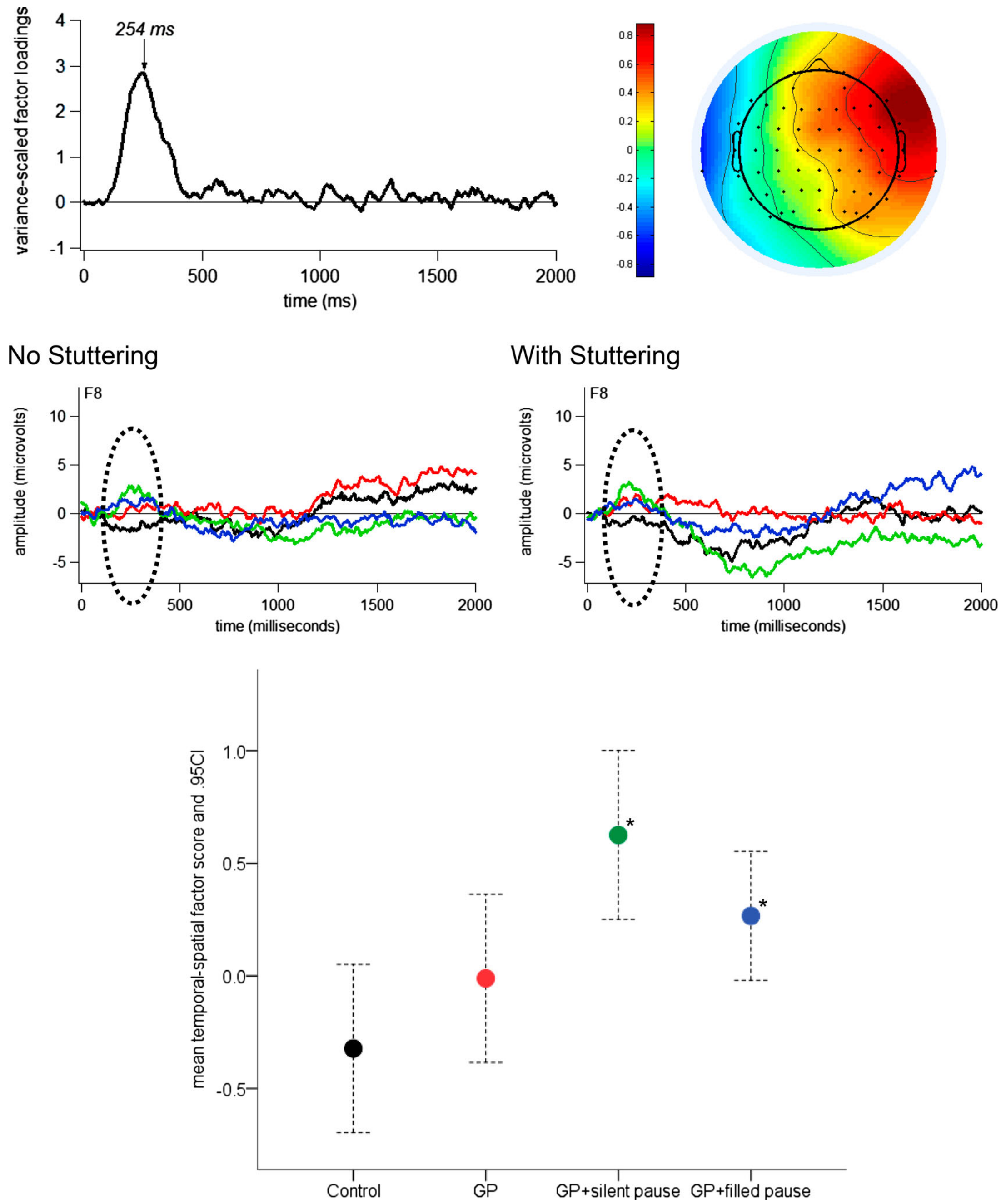

Figure 4. Factor loadings for T254 (top left). Topographic map of the right anterior spatial factor associated with T254 (top right). Factor scores summarising the ERP variance within the T254 time window at this scalp region (bottom) (black = Control, red = GP, green = GP + silent pause, blue = GP + filled pause). Asterisks indicate conditions that differed in amplitude from Control ($p < .016$). Illustration of this component activity in grand average waveforms for each group at electrode F8 (middle).

spatial factor scores – summarising ERP activity at ~648 ms after verb onset at the posterior scalp region – were affected by Sentence Type ($F[3,84] = 4.24$, $p = .006$) and an interaction of Group and Sentence Type ($F[3,84] = 3.12$, $p = .03$). Bonferroni-corrected *t*-tests revealed that, for the No Stuttering group only, T648/posterior activity to GP was positive-going versus Control ($p = .005$) (see Figure 6). Positive-going activity to the GP condition is visually evident at ~650 ms in the grand average of the No Stuttering group (e.g. at electrode PO4 in Figure 2). This effect was not detected for GP + silent pause ($p = .27$) or GP + filled pause ($p = .21$) versus Control in the No Stuttering group. Nor was this effect detected for GP ($p = .81$), GP + silent pause ($p = .79$), or GP + filled pause ($p = .93$) versus Control in the With Stuttering group.

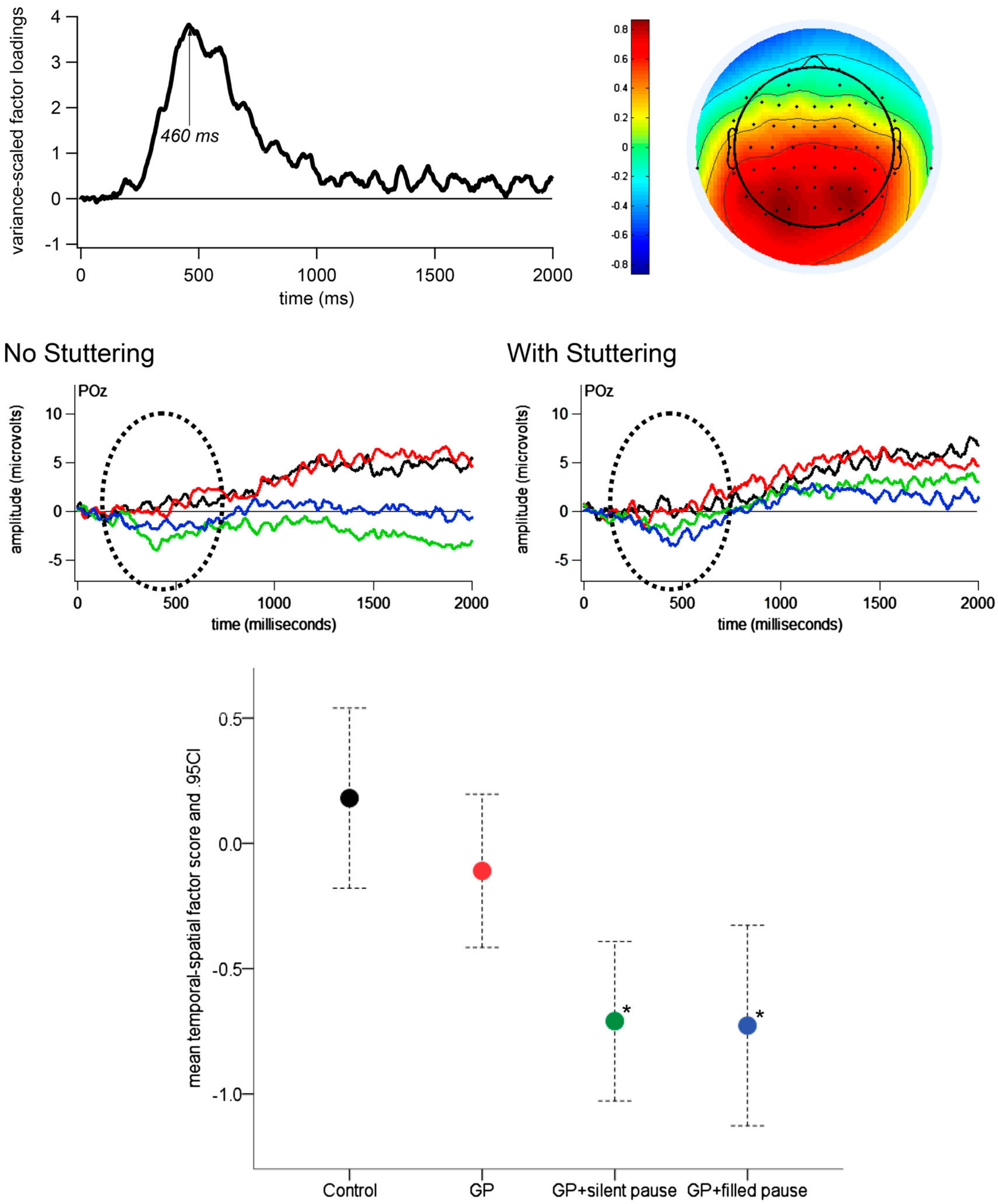

Figure 5. Factor loadings for T460 (top left). Topographic map of the posterior spatial factor associated with T460 (top right). Factor scores summarising the ERP variance within the T460 time window at this scalp region (bottom) (black = Control, red = GP, green = GP + silent pause, blue = GP + filled pause). Asterisks indicate conditions that differed in amplitude from Control ($p < .016$). Illustration of this component activity in grand average waveforms for each group at electrode POz (middle).

T1196, left anterior activity

Three spatial factors were Promax-rotated for T1196, partitioning the voltage variance in this time window into three scalp regions. One spatial factor had a left anterior scalp topography (see Figure 7). Temporal-spatial factor scores – summarising ERP activity at ~1196 ms after verb onset at the left anterior scalp region – were affected by Sentence Type ($F[3,84] = 4.95$, $p = .003$). Bonferroni-corrected *t*-tests revealed that positive-going activity elicited to Control was attenuated to GP + silent pause ($p = .002$) (see Figure 7). This same effect approached statistical significance for Control versus GP + filled pause ($p = .03$). Positive-going activity to the Control (and GP) condition is visually evident at ~1200 ms in the grand averages of both groups (e.g. at electrode F7 in Figures 2 and 3).

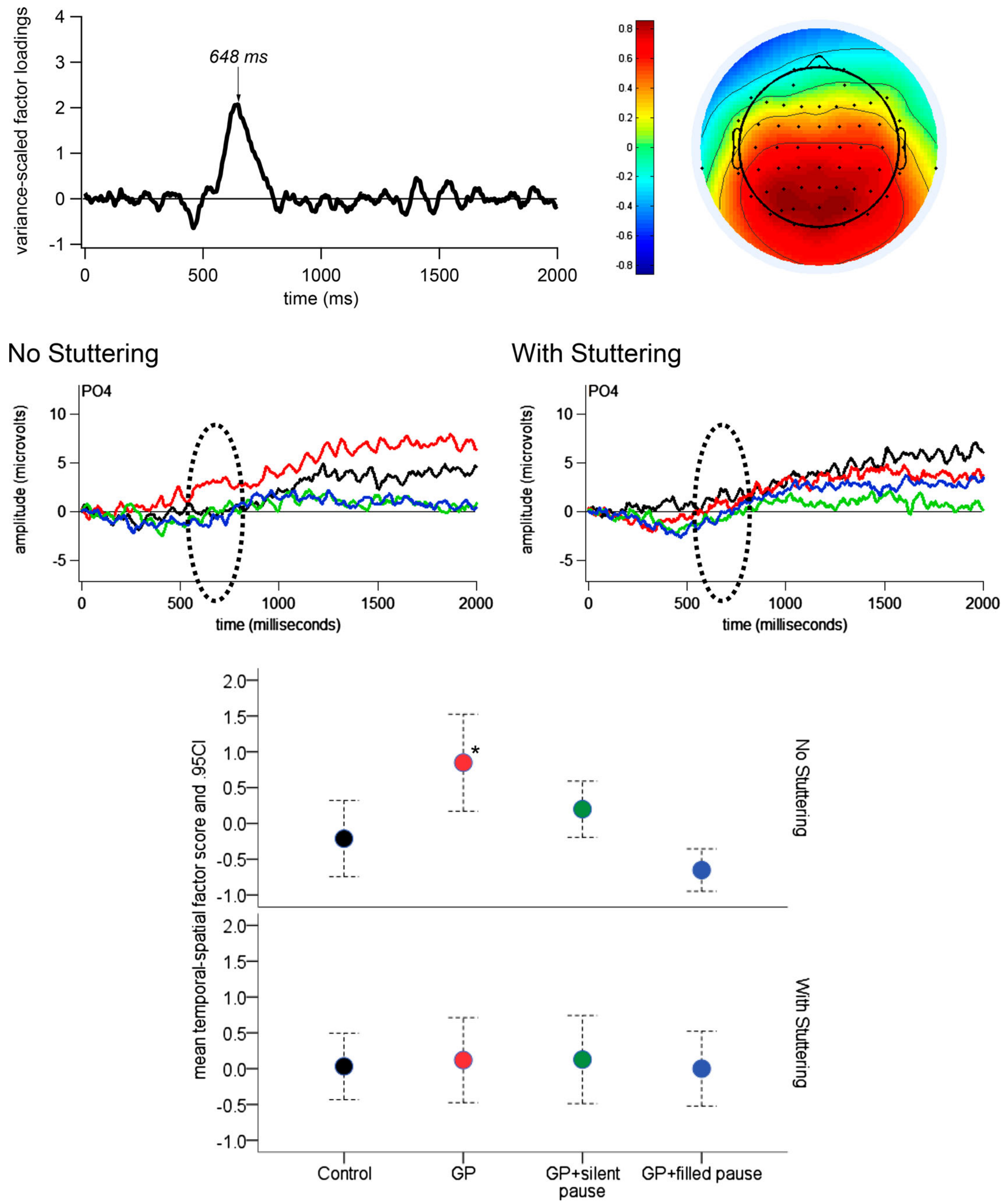

Figure 6. Factor loadings for T648 (top left). Topographic map of the posterior spatial factor associated with T648 (top right). Factor scores summarising the ERP variance within the T648 time window at this scalp region (bottom) (black = Control, red = GP, green = GP + silent pause, blue = GP + filled pause). Asterisk indicates a condition that differed in amplitude from Control ($p < .016$). Illustration of this component activity in grand average waveforms for each group at electrode PO4 (middle).

Discussion

Two groups of listeners heard a set of GP and control sentences. Some of the GP sentences contained a typical disfluency, in the form of a silent or filled pause, just before the disambiguating verb. The sentences presented to each group were produced by the same speaker, an adult male with a diagnosis of persistent developmental stuttering who, following years of speech therapy, was able to read aloud in the absence of stuttering using controlled fluency. One group of listeners heard the sentences produced with controlled fluency (No Stuttering), while the other group of listeners heard the same sentences but containing some stuttering-like disfluencies (With Stuttering), which the speaker produced by turning off fluency controls and allowing himself to stutter openly once or twice in the sentence preamble. Listeners in the No Stuttering group more

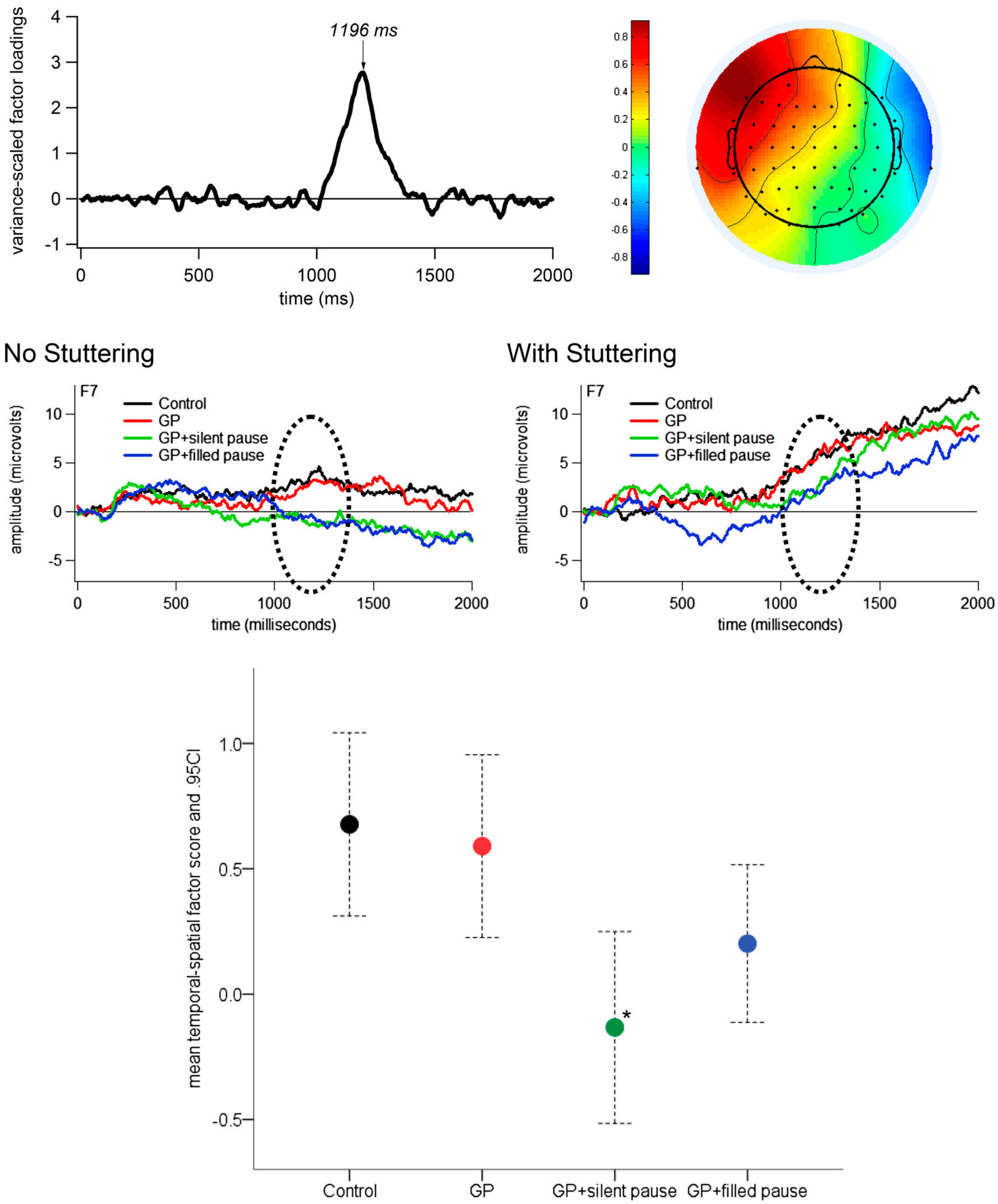

Figure 7. Factor loadings for T1196 (top left). Topographic map of the left anterior spatial factor associated with T1196 (top right). Factor scores summarising the ERP variance within the T1196 time window at this scalp region (bottom) (black = Control, red = GP, green = GP + silent pause, blue = GP + filled pause). Asterisk indicates a condition that differed in amplitude from Control ($p < .016$). Illustration of this component activity in grand average waveforms for each group at electrode F7 (middle).

often agreed that the speaker was competent, fluent, read the sentences easily, and was comfortable to listen to, than listeners in the With Stuttering group.

A probe question, presented after each sentence, assessed listeners' interpretations of either the subordinate or matrix clause. In general, comprehension was less accurate for GP versus Control sentences in each group, although accuracy to GP sentences was better when asked about matrix versus subordinate clauses. Crucially, comprehension of GP, GP + silent pause and GP + filled pause sentences was more accurate in the With Stuttering group than in the No Stuttering group. Additionally, comprehension accuracy in the No Stuttering group was moderated by listener perceptions of the speaker.

Scalp-recorded ERP activity, elicited to the critical verb in each sentence, included a P600 component to GP sentences but not to GP + silent pause or GP + filled pause sentences in the No Stuttering group. In contrast, P600 was not detected to GP, GP + silent pause or GP + filled pause sentences in the With Stuttering group. Other ERP components elicited to GP + silent pause and GP + filled pause sentences (i.e. an early right anterior/temporal positivity, a posterior N400 effect, and a late left anterior effect) were not found to differ in their presence, timing or scalp topography between groups.

Comprehension accuracy

Probe question accuracy was affected by Sentence Type, Clause Type and their interaction. The observed effects mirror those reported in previous research (e.g. Bailey & Ferreira, 2003; Christianson et al., 2001; Maxfield et al., 2009). Poorer accuracy to subordinate clause questions (e.g. *Did the man hunt the deer?*, following the sentence *While Bill hunted the deer ran into the woods*) than to matrix clause questions (e.g. *Did the deer run away?*) about GP sentences, suggests that listeners sometimes arrived at, and committed to, initially plausible albeit incorrect interpretations of those sentences. Probe question accuracy to GP sentences was even worse when the ambiguous region was lengthened by the presence of silent and filled pauses, suggesting that "lingering" on a plausible interpretation made listeners even more likely to commit to it even though disambiguating information followed.

Consistent with a "good enough" model of sentence processing (Ferreira, Bailey, & Ferraro, 2002), the pattern of results observed here suggests that listeners may use a combination of plausibility and timeliness of information to interpret sentences on a rapid and shallow basis, even though this strategy can result in incorrect interpretations. Of interest, the more strongly listeners in the No Stuttering group agreed the speaker was fluent the better the accuracy with which they interpreted the subordinate clause in GP + filled pause sentences as well as the matrix clause in all three types of GP sentences. This suggests that the more confident listeners were in the speaker, the more likely they were to correctly interpret the second NP in each sentence as the subject/agent of the matrix clause and, at least in the case of GP + filled pause sentences, drop the incorrect interpretation of the second NP as the object/theme of the subordinate clause.

On the other hand, "good enough" processing seemed less pronounced in the With Stuttering group. Accuracy for this group was better than the No Stuttering group to GP, GP + silent pause and GP + filled pause sentences. As discussed next, ERP results together with probe question accuracy, suggest that listeners processing sentences With Stuttering were less likely to form initially erroneous interpretations based on the plausibility of information appearing early in sentences.

P600 effects

For the No Stuttering group, we replicated the finding by Maxfield et al. (2009) that GP sentences elicited a P600 effect while this effect was attenuated or absent (i.e. statistically undetectable) to GP + silent pause or GP + filled pause sentences. As discussed in that paper, the P600 component has been interpreted as an ERP correlate of parse revision, elicited during sentence processing when an inappropriate, less-preferred or ambiguous syntactic structure is encountered (Frisch et al., 2002). That this effect was undetectable to GP + silent pause or to GP + filled pause sentences in the No Stuttering group suggests that lengthening the ambiguous region in GP sentences with a disfluency interrupts the parse revision process indexed by P600, a possibility first entertained by Christianson et al. (2001) and Bailey and Ferreira (2003).

For the With Stuttering group, a P600 effect was not detected to GP, GP + silent pause or GP + filled pause sentences. As discussed previously, an undetectable P600 effect to GP + silent pause and GP + filled pause sentences can be interpreted as suggesting that when listeners linger on an initial sentence interpretation they may commit to that interpretation even when disambiguating information follows (i.e. P600-indexed revision goes undetected). From this perspective, an undetectable P600 effect to GP sentences might suggest that listeners in the With Stuttering group also formed initial (erroneous) sentence interpretations that were not later revised at the disambiguating verb, possibly because the presence of stuttering-like disfluencies in the sentence preambles forced them to linger on, and thereby commit to, initial interpretations. If so, then probe question accuracy to GP sentences might be expected to be reduced for the With Stuttering versus No Stuttering group. However, the opposite was observed, with probe question accuracy better for the With Stuttering group to GP, GP + silent pause and GP + filled pause sentences. This suggests that, in all three GP conditions, listeners in the With Stuttering group refrained from forming, or committing to, interpretations based on information contained in the sentence preambles, thereby arriving at correct interpretations more often than the No Stuttering group (and reducing P600-indexed revisions to GP sentences in the With Stuttering group).

Right anterior/temporal positivity

Beyond P600, other ERP components were modulated by Sentence Type. One was a right anterior/temporal positivity to disambiguating verbs appearing in both GP +

silent pause and GP + filled pause sentences. This effect was not differentiated by Group, or by an interaction of Group and Sentence Type. In Maxfield et al. (2009), a similar effect detected to the same sentence conditions appeared at approximately the same latency but had a left rather than right anterior/temporal scalp topography. We tentatively related that effect to a left anterior positivity observed by Posner and colleagues (e.g. Posner & Pavese, 1998) to isolated words.

Other anterior positivities have been reported in language processing too. Holcomb, Coffey, and Neville (1992) observed an anterior P250 component to auditory words completing sentences. The amplitude of that P250 component was attenuated to anomalous versus best completions. Still other right-hemisphere ERP effects have been observed in processing of figurative language (Coulson & Van Petten, 2007). Additional research will be necessary to determine whether any of these effects relate to right anterior/temporal positive-going activity elicited here to disambiguating verbs in GP + silent pause and GP + filled pause sentences. Whatever process is associated with this positive-going activity, it was not significantly affected by the presence versus absence of stuttering-like disfluencies.

N400 effect

Disambiguating verbs in GP + silent pause and GP + filled pause sentences also elicited a posterior N400 effect. These same sentence conditions also elicited posterior N400 in Maxfield et al. (2009). As discussed there, N400 marks an attempt to integrate new information into the current interpretation of a sentence, particularly when the listener "trusts" that sentence interpretation (Kolk, Chwilla, Van Herten, & Oor, 2003).

From that perspective, it seems noteworthy that N400 effects did not differ between groups. As reviewed in the Introduction, listeners may perceive speech produced by people who stutter not only as more effortful to process but also as distracting and difficult to recall. Despite those perceptions of stuttered speech, the current results suggest that attempts to integrate disambiguating verbs into current sentence interpretations, following silent and filled pauses, were no less likely when processing sentences With Stuttering than when processing No Stuttering sentences.

Late effect

Finally, a late ERP effect differentiated processing of GP + silent pause from processing of Control sentences. This effect was characterised by left anterior positive-going activity to Control sentences that was attenuated to GP + silent pause sentences and (marginally) to GP + filled pause sentences. GP sentences without silent or filled pauses also elicited a late left anterior positivity.

Slow positive drifts at the left anterior scalp region have been linked to working memory operations involved with forming a mental model (Kutas & King, 1996). Applying this interpretation to left anterior positivity-going effects here would suggest that – as listeners approached the ends of Control and GP sentences – syntactic structures, word meanings and other sources of information that were analysed and held in working memory as each sentence was presented were integrated to set-up a final sentence interpretation. The presence of silent and filled pauses appears to have disrupted this activity. Noteworthy too was the finding that late left anterior positive-going activity to Control and GP sentences did not differ between groups, suggesting that high-level integration in processing of Control and GP sentences was not affected by the presence versus absence of stuttering.

Listener perceptions

All but one of the listeners in the With Stuttering group described the stimuli using the word "stutter". It seems noteworthy that a subset of listeners in the No Stuttering group also used the word "stutter" to characterise the stimuli. Unknown is whether listeners in the No Stuttering group used the term "stutter" to characterise the frequent occurrence of filled and unfilled pauses (the term "stutter" is commonly used to characterise speech that does not necessarily contain stuttering-like disfluencies), or whether characteristics of the No Stuttering sentences contributed to perceptions that the speech was produced by a person diagnosed with stuttering. As noted in the Introduction, listeners may perceive speech produced by people who stutter as different even in the absence of explicit stuttering (Susca & Healey, 2001).

Overall, it is important to emphasise that listeners in the No Stuttering group more often agreed the speaker was fluent and read the sentences easily than listeners in the With Stuttering group. This is consistent with previous research outlined in the Introduction, and confirms that listeners in our two groups had different impressions of the same speaker depending on the presence of controlled fluency versus stuttering-like disfluencies. It seems noteworthy that listeners in the No Stuttering group also more often agreed the speaker was competent and comfortable to listen to. Listeners in other studies preferred mildly stuttered speech over controlled fluency (e.g. Manning, Burlison, &

Thaxton, 1999). In contrast, our speaker seems to have managed fluency with enough naturalness to elicit more favourable impressions with controlled fluency than With Stuttering.

Summary and conclusions

The current results suggest that in the presence of stuttering-like disfluencies, listeners may refrain from forming, or committing to, sentence interpretations based on information appearing early in sentences. This strategy seems more consistent with the Noisy Channel model of sentence processing, although not necessarily in the manner suggested in the Introduction. As outlined there, one hypothesis was that listeners might assign a different status to silent and filled pauses appearing in sentences. With Stuttering which might, in turn, result in P600-indexed sentence revisions to GP, GP + silent pause and GP + filled pause sentences produced With Stuttering. Instead, P600 was undetectable to any of the GP sentences produced With Stuttering, but accuracy was better than when those same sentences were produced with No Stuttering, suggesting that listeners adapted to sentences With Stuttering by moving away from "good enough" sentence processing.

One way to think about this shift away from "good enough" processing blends that idea with the Noisy Channel approach and invokes the idea of speaker modelling. If "good enough" processing is interpreted as due to listeners being overly reliant on expectations ("priors"; Kuperberg & Jaeger, 2016) at the expense of the data (i.e. the input), then these results might suggest that listeners are reluctant to use their priors when presented with speech that is different from that from which most of those expectations were derived. This would lead listeners to focus on the input itself and perhaps downplay any expectations, allowing them to avoid premature commitments to any syntactic structure (whether filled or unfilled pauses were present or not), and thus reducing the likelihood of experiencing a GP. As a result, no P600 would be observed, and probe accuracy would be higher than to No Stuttering sentences. The argument, then, is that "good enough" processing is less likely in the With Stuttering condition because the listener has limited experience with such speech and thus has an incomplete or sketchy speaker model, leading the listener to rely on the data rather than on expectations based on experience.

Several other ERP effects (i.e. early positive-going activity, later N400 activity, and even later frontal positive-going activity) were not affected by the presence versus absence of stuttering-like disfluencies. This seems more consistent with the Ambiguity Resolution model of sentence processing which, as outlined in the Introduction, claims that sentence processing proceeds regardless of speaker identity. One implication of these findings is that some aspects of sentence processing (e.g. "good enough" processing) are vulnerable to the presence of stuttering-like disfluencies while others are not.

Another implication of the current findings is that relatively small amounts of stuttering-like disfluency can trigger more thorough processing of sentences containing ambiguities. Additional research is necessary to determine whether this type of processing is more versus less taxing on cognitive resources and whether any cognitive trade-offs benefit listener performance. Additional research is also necessary to identify whether there is a "threshold" at which the frequency, type and/or location of stuttering produces more negative effects on processing of ambiguous sentences. It is also important to continue investigating whether stuttering-like disfluencies impact other aspects of sentence processing. Findings from this line of research may, ultimately, have implications for guiding therapeutic interventions focused on improving communication effectiveness of people who stutter.

Notes

1. We acknowledge that speakers who stutter do not necessarily stutter on every utterance produced, although some might. For purposes of this study, stuttering was included in each sentence heard by the With Stuttering group to ensure that listeners in this group would unambiguously identify the speaker as someone who stutters.
2. It is important to note that sentence interpretations and, thus, responses to probe questions may be driven by pragmatic and syntactic processes operating in tandem (Bailey & Ferreira, 2003). Thus, in response to the GP sentence, "While the man hunted the deer ran into the woods", listeners may answer *Yes* when asked "Did the man hunt the deer?" based on the pragmatic inference that the hunter was hunting the deer, even though syntactically the deer is not the object of the hunter. For all of the sentences used in this experiment, the first and second noun phrases were plausibly related, allowing for this same kind of pragmatic inference to be made. Thus, responses to probe questions were not treated as wholly correct or incorrect but, rather, as reflecting listener interpretations based on pragmatics and/or syntax. For this reason, all trials (whether interpreted correctly from a syntactic perspective or not) were included in the analysis of the ERP data.
3. For Control and GP sentences, the critical verb was preceded immediately by lexical material. For GP-disfluency

sentences, the critical verb was preceded immediately by a filled or silent pause. Our concern was that the different material preceding the critical verbs in the different conditions elicited different types of ERP activity, unevenly affecting ERP amplitude immediately preceding and at the onset of presentation of our critical verbs in the different conditions. Other published studies have used a post-stimulus baseline correction procedure to control for this possibility (e.g. Neville, Nichol, Barss, Forster, & Garrett, 1991; Friederici, Hahne, & Mecklinger, 1996; Hahne, 2001; Hahne & Friederici, 1999; Hahne & Friederici, 2001; Osterhout et al., 1994; Maxfield et al., 2009). We adopted their approach. In order to validate this procedure, Friederici et al. (1996) (among other examples) first baseline-corrected their averaged ERPs using a post-stimulus interval. They then determined, via statistical analysis, that amplitude differences were not present in the ERPs between conditions during the post-stimulus baseline interval. Along this same line, our Principal Component Analysis (reported below) failed to uncover any statistically significant ERP amplitude differences between conditions in the duration covering our post-stimulus baseline interval (0 to +100 ms), i.e. there were no temporal factors with a peak latency between 0 and +100 ms post-verb that captured ERP amplitude differences between conditions.

Acknowledgements

We are grateful to Alissa Belmont for her efforts in preparing the stimuli, collecting and processing the data presented here. We are also grateful to the speaker who stutters who provided verbal stimuli for this study.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was supported by the National Institutes of Health – National Institute of Deafness and other Communication Disorders (NIH-NIDCD) under grant number (R56DC013545). The first author was also supported by NIH-NIDCD grant (R03DC011144) during work on this study.

References

Bailey, K. G., & Ferreira, F. (2003). Disfluencies affect the parsing of garden-path sentences. *Journal of Memory and Language*, *49*(2), 183–200.

Beilby, J. M., Byrnes, M. L., Meagher, E. L., & Yaruss, J. S. (2013). The impact of stuttering on adults who stutter and their partners. *Journal of Fluency Disorders*, *38*(1), 14–29.

Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, *7*(6), 1129–1159.

Blomgren, M. (2010). Stuttering treatment for adults: An update on contemporary approaches. *Seminars in Speech and Language*, *31*(4), 272–282.

Brown, S. F. (1945). The loci of stutterings in the speech sequence. *Journal of Speech Disorders*, *10*(3), 181–192.

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, *42*(4), 368–407.

Coulson, S., & Van Petten, C. (2007). A special role for the right hemisphere in metaphor comprehension?: ERP evidence from hemifield presentation. *Brain Research*, *1146*, 128–145.

Craig, A., Blumgart, E., & Tran, Y. (2009). The impact of stuttering on the quality of life in adults who stutter. *Journal of Fluency Disorders*, *34*(2), 61–71.

Cyprus, S., Hezel, R. T., Rossi, D., & Adams, M. R. (1984). Effects of simulated stuttering on listener recall. *Journal of Fluency Disorders*, *9*(3), 191–197.

Dien, J. (2010). The ERP PCA toolkit: An open source program for advanced statistical analysis of event-related potential data. *Journal of Neuroscience Methods*, *187*(1), 138–145.

Dien, J., & Frishkoff, G. A. (2005). Principal components analysis of event-related potential datasets. In T. Handy (Ed.), *Event-related potentials: A methods handbook* (pp. 189–208). Cambridge: MIT Press.

Ferreira, F., Bailey, K. G., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*(1), 11–15.

Friederici, A. D. (1998). Diagnosis and reanalysis: Two processing aspects the brain may differentiate. In J. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 177–200). Dordrecht: Springer.

Friederici, A. D., Hahne, A., & Mecklinger, A. (1996). Temporal structure of syntactic parsing: Early and late event-related brain potential effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1219.

Frisch, S., Schlesewsky, M., Saddy, D., & Alpermann, A. (2002). The P600 as an indicator of syntactic ambiguity. *Cognition*, *85*(3), B83–B92.

Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*, *110*(20), 8051–8056.

Glass, K. A., Frishkoff, G. A., Frank, R. M., Davey, C., Dien, J., Malony, A. D., & Tucker, D. M. (2004). A framework for evaluating ICA methods of artifact removal from multichannel EEG. In *International conference on independent component analysis and signal separation* (pp. 1033–1040). Berlin: Springer.

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*(2), 95–112.

Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis: Early automatic and late controlled processes. *Journal of Cognitive Neuroscience*, *11*(2), 194–205.

Hahne, A. (2001). What's different in second-language processing? Evidence from event-related brain potentials. *Journal of Psycholinguistic Research*, *30*(3), 251–266.

Hahne, A., & Friederici, A. D. (2001). Processing a second language: Late learners' comprehension mechanisms as revealed by event-related brain potentials. *Bilingualism: Language and Cognition*, *4*(2), 123–141.

Hannah, E. P., & Gardner, J. G. (1968). A note on syntactic relationships in nonfluency. *Journal of Speech, Language, and Hearing Research*, *11*(4), 853–860.

Hanulíková, A., Van Alphen, P. M., Van Goch, M. M., & Weber, A. (2012). When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience*, *24*(4), 878–887.

Healey, E. C. (2010). What the literature tells us about listeners' reactions to stuttering: Implications for the clinical management of stuttering. *Seminars in Speech and Language*, *31*(4), 227–235.

Hendrickson, A. E., & White, P. O. (1964). Promax: A quick method for rotation to oblique simple structure. *British Journal of Statistical Psychology*, *17*(1), 65–70.

Holcomb, P. J., Coffey, S. A., & Neville, H. J. (1992). Visual and auditory sentence processing: A developmental analysis using event-related brain potentials. *Developmental Neuropsychology*, *8*(2-3), 203–241.

Howell, P., Au-Yeung, J., & Sackin, S. (1999). Exchange of stuttering from function words to content words with age. *Journal of Speech, Language, and Hearing Research*, *42*(2), 345–354.

Hulit, L. M. (1976). Effects of nonfluencies on comprehension. *Perceptual and Motor Skills*, *42*(3_suppl), 1119–1122.

Iverach, L., O'Brian, S., Jones, M., Block, S., Lincoln, M., Harrison, E., … Onslow, M. (2009). Prevalence of anxiety disorders among adults seeking speech therapy for stuttering. *Journal of Anxiety Disorders*, *23*(7), 928–934.

Jayaram, M. (1984). Distribution of stuttering in sentences: Relationship to sentence length and clause position. *Journal of Speech and Hearing Research*, *27*(3), 338–341.

Kaasin, K., & Bjerkan, B. (1982). Critical words and the locus of stuttering in speech. *Journal of Fluency Disorders*, *7*(4), 433–446.

Klem, G. H., Lüders, H. O., Jasper, H. H., & Elger, C. (1999). The ten-twenty electrode system of the International Federation. *Electroencephalography and Clinical Neurophysiology*, *52*(3), 3–6.

Kolk, H. H., Chwilla, D. J., Van Herten, M., & Oor, P. J. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language*, *85*(1), 1–36.

Koopmans, M., Slis, I., & Rietveld, T. (1991). The influence of word position and word type on the incidence of stuttering. In H. F. M. Peters, W. Hulstijn, & C. W. Starkweather (Eds.), *Speech motor control and stuttering* (pp. 333–340). New York: Elsevier.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647.

Kutas, M., & King, J. W. (1996). The potentials for basic sentence processing: Differentiating integrative processes. In I. Ikeda & J. L. McClelland (Eds.), *Attention and performance* (pp. 501–546). Cambridge: MIT Press.

Lanyon, R. I. (1969). Behavior change in stuttering through systematic desensitization. *Journal of Speech and Hearing Disorders*, *34*(3), 253–260.

Logan, K. J. (2003). The effect of syntactic structure upon speech initiation times of stuttering and nonstuttering speakers. *Journal of Fluency Disorders*, *28*(1), 17–35.

Lowder, M., Maxfield, N. D., & Ferreira, F. (submitted). Processing of self-repairs in stuttered and non-stuttered speech.

Manning, W. H., Burlison, A. E., & Thaxton, D. (1999). Listener response to stuttering modification techniques. *Journal of Fluency Disorders*, *24*(4), 267–280.

Maxfield, N. D., Lyon, J. M., & Silliman, E. R. (2009). Disfluencies along the garden path: Brain electrophysiological evidence of disrupted sentence processing. *Brain and Language*, *111*(2), 86–100.

Neville, H., Nicol, J. L., Barss, A., Forster, K. I., & Garrett, M. F. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*, *3*(2), 151–165.

Nunez, P. L., & Srinivasan, R. (2006). *Electric fields of the brain: The neurophysics of EEG*. New York: Oxford University Press.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113.

Osterhout, L., Holcomb, P. J., & Swinney, D. A. (1994). Brain potentials elicited by garden-path sentences: Evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 786.

Panico, J., & Healey, E. C. (2009). Influence of text type, topic familiarity, and stuttering frequency on listener recall, comprehension, and mental effort. *Journal of Speech, Language, and Hearing Research*, *52*(2), 534–546.

Panico, J., Healey, E. C., Brouwer, K., & Susca, M. (2005). Listener perceptions of stuttering across two presentation modes: A quantitative and qualitative approach. *Journal of Fluency Disorders*, *30*(1), 65–85.

Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson, R., … Taylor, M. J. (2000). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology*, *37*(2), 127–152.

Posner, M. I., & Pavese, A. (1998). Anatomy of word and sentence meaning. *Proceedings of the National Academy of Sciences*, *95*(3), 899–905.

Preisendorfer, R. W., & Mobley, C. D. (1988). *Principal component analysis in meteorology and oceanography* (Vol. 17). Amsterdam: Elsevier.

Prins, D., & Ingham, R. J. (2009). Evidence-based treatment and stuttering—Historical perspective. *Journal of Speech, Language, and Hearing Research*, *52*(1), 254–263.

Quarrington, B. (1956). Cyclical variation in stuttering frequency and some related forms of variation. *Canadian Journal of Psychology*, *10*(3), 179–184.

Ratner, N. B., & Benitez, M. (1985). Linguistic analysis of a bilingual stutterer. *Journal of Fluency Disorders*, *10*(3), 211–219.

Richman, M. B. (1986). Rotation of principal components. *Journal of Climatology*, *6*(3), 293–335.

Sander, E. K. (1965). Comments on investigating listener reaction to speech disfluency. *Journal of Speech & Hearing Disorders*, *30*, 159–165.

Silverman, S. W., & Ratner, N. B. (1997). Syntactic complexity, fluency, and accuracy of sentence imitation in adolescents. *Journal of Speech, Language, and Hearing Research*, *40*(1), 95–106.

Soderberg, G. A. (1971). Relations of word information and word length to stuttering disfluencies. *Journal of Communication Disorders*, *4*(1), 9–14.

Susca, M., & Healey, E. C. (2001). Perceptions of simulated stuttering and fluency. *Journal of Speech, Language, and Hearing Research*, *44*(1), 61–72.

Susca, M., & Healey, E. C. (2002). Listener perceptions along a fluency–disfluency continuum: A phenomenological analysis. *Journal of Fluency Disorders*, *27*(2), 135–161.

Tataryn, D. J., Wood, J. M., & Gorsuch, R. L. (1999). Setting the value of k in promax: A Monte Carlo study. *Educational and Psychological Measurement*, *59*(3), 384–391.

Tornick, G. B., & Bloodstein, O. (1976). Stuttering and sentence length. *Journal of Speech, Language, and Hearing Research*, *19*(4), 651–654.

Tsiamtsiouris, J., & Cairns, H. S. (2013). Effects of sentence-structure complexity on speech initiation time and disfluency. *Journal of Fluency Disorders*, *38*(1), 30–44.

Wells, G. B. (1979). Effect of sentence structure on stuttering. *Journal of Fluency Disorders*, *4*(2), 123–129.

Wingate, M. E. (1964). A standard definition of stuttering. *Journal of Speech and Hearing Disorders*, *29*(4), 484–489.

Yaruss, J. S. (2010). Assessing quality of life in stuttering treatment outcomes research. *Journal of Fluency Disorders*, *35*(3), 190–202.