Testing predictions about the processing of word stress in reading using event-related potentials

Conrad Perry

To cite this article: Conrad Perry (2017): Testing predictions about the processing of word stress in reading using event-related potentials, Language, Cognition and Neuroscience, DOI: 10.1080/23273798.2017.1398343

To link to this article: http://dx.doi.org/10.1080/23273798.2017.1398343

Published online: 07 Nov 2017.
Testing predictions about the processing of word stress in reading using event-related potentials

Conrad Perry
Faculty of Health, Arts and Design, Swinburne University of Technology, Hawthorn, Australia

ABSTRACT
Both computational models of English reading that generate word stress predict a processing advantage for words with initial syllable stress. They differ, however, on whether they process words incrementally and learn nonlinear spelling-stress relationships. Two experiments using event-related potentials were used to investigate these predictions. The first examined trisyllabic stimuli. Differences found on P200 and N400 components suggested a processing advantage for words with initial syllable stress. The second examined root morphemes within words that have high frequency suffixes that are stress predictive. A processing advantage on the N400 component was found with root morphemes that typically have initial syllable stress, even when the whole-word stress pattern differed. This provides evidence that stress is generated incrementally, where it is assigned to parts of words as they are processed, and that stress assignment is not necessarily affected by high frequency nonlinear relationships.

One way of trying to understand how word stress is generated when reading is to look at predictions that computational models of reading make. At present, there are only two computational models that deal with word stress in English, CDP++ (Perry, Ziegler, & Zorzi, 2010, 2013; see Figure 1) and the PDP model of Ševa, Monaghan, and Arciuli (2009) (see also Arciuli, Monaghan, & Seva, 2010). A complex algorithm devised by Rastle and Coltheart (2000) also exists, although it was not proposed as a model of the psychological mechanisms used. Both of these models and the algorithm were only designed to process at most disyllabic words, although the principles that underlie them can be used to help understand how more complex forms of stress might be processed.

CDP++ is a full model of reading aloud, and is able to generate both segmental phonology and word stress. Both of these can be generated two ways, one via a lexical method where the phonemes and the stress pattern of a word are simply looked up from a stored lexical form. The other via a sublexical method where phonemes and stress are imputed from the letters without whole-word knowledge. With the sublexical method, letters are first grouped into multi-letter graphemes and these are placed in a syllabic template via the use of a graphemic parser. This is done incrementally, one grapheme at a time. A simple two-layer linear network imputes phonology (segmental and stress) from them. The results from this and the lexical route are then combined in a phonological and a stress output buffer. The stress output buffer has two nodes, one for each syllable. Effects of spelling-stress consistency, which is the extent to which the letters of a word are able to predict the stress pattern of the word, may arise at this level because the phonology generated from lexical and sublexical methods may differ and cause competition before the final form is chosen. Similarly, effects of spelling-segmental phonology regularity, which is the extent to which the letters of a word are able to predict the segmental phonology, may also arise for the same reason (this is generally known as spelling-sound consistency or spelling-sound regularity, but spelling-segmental phonology regularity is used here to avoid confusion).

One of the properties of the sublexical route of CDP++ that is in part responsible for its good performance compared to other models (e.g. Perry et al., 2010) is that it uses a simple linear network trained with the Rescorla-Wagner rule (e.g. Sutton & Barto, 1981) to learn regularities between letters and stress patterns. This allows the model to learn quickly and learning is relatively stable. Interestingly, after extensive testing (Perry et al., 2010, 2013), the sublexical route shows some bias towards the most common pattern where the initial syllable of nonwords is stressed more often than the final one. That is, the sublexical stress route tends to default to a trochaic pattern (stressed-unstressed) if there is no information learnt from the letters to suggest an iambic pattern
should be used. Thus, the model predicts that words that start with initial stress will be processed differently to those that do not because they tend to benefit more from sublexical information.

The model of Ševa et al. (2009) is a 3-layer backpropagation style network and is similar to CDP++ in that it learns to generalise to stress based on spelling-sound regularities. It also predicts that there is a bias towards initial over second syllable stress. However, it differs in a number of important ways. One is that all letters of a word are presented to it in parallel. This means that it generates stress as soon as a word is presented to it. Thus, there is no opportunity for the letters to be presented incrementally, unlike CDP++. A second difference is that it allows nonlinear relationships between spelling and stress to be learnt. This means that sublexical stress can be generated correctly for any word, and if reliable and frequent cues are present, the amount of error the model will make when generating stress will be minimised.

**Behavioural data**

The results of experiments examining the effect of spelling-stress consistency have been mixed. In English, the most common pattern found in behavioural tasks is an effect on error rates (e.g. Arciuli & Cupples, 2003, 2006; Arciuli & Monaghan, 2009; Chateau & Jared, 2003; Yap & Balota, 2009). However, some experiments have found no effect at all (Rastle & Coltheart, 2000), and an effect on reaction times has only been reported sporadically (Arciuli & Monaghan, 2009; Mundy & Carroll, 2012; Rastle & Coltheart, 2000). In other languages like Italian where stress has been extensively investigated, effects have generally been found both on reaction times and error rates (see Sulpizio, Burani, & Colombo, 2015, for a review).

One reason why mixed results of spelling-stress consistency may have been found in English is because there is no simple definition of what spelling-stress consistency actually is. Studies have used definitions based on whether stress simply occurs on the initial or second syllable (e.g. Chateau & Jared, 2003; Rastle & Coltheart, 2000; Yap & Balota, 2009), and more complex algorithms (e.g. Rastle & Coltheart, 2000). Even using a definition just based on the final letters in a word, which works well for Italian (e.g. Sulpizio, Burani, et al., 2015), is not perfect in English because there is no simple
delineation of potential structural groups that can be used in words, and something that is consistent on one level may be inconsistent on another (e.g. Fudge, 1984). For example, the word carbonic (second syllable stress) is consistent based on its final letters (-ic) because most words that end with -ic in the ultimate syllable generally have stress on the penultimate one, but it is inconsistent because most words that use carbon as a root morpheme do not have stress on the second syllable (e.g. carbonated, carbonless).

A second reason that effects of spelling-stress consistency may have been difficult to find is that some types of behavioural tasks are susceptible to effects that would obscure them. Notably, people may generate a response rhythm that causes them to respond within certain bounds (e.g. Lupker, Brown, & Colombo, 1997) and speech output processes may also obscure some effects (e.g. Sulpizio, Spinelli, & Burani, 2015). Thus, people may respond before stress has been correctly resolved and thus differences are often found on error rates and not reaction times. Such differences suggest that the time-course of stress generation may differ compared to other phonological effects where differences in both reaction times and error rates are usually found, such as when spelling-segmental phonology regularity has been examined in tasks that require speech output (e.g. Jared, 2002; Ziegler, Perry, & Coltheart, 2003).

Electroencephalography data

Given the complications that speech output may cause when examining stress, electroencephalography (EEG) may offer insight into the processing of stress as the time-course of processing can be examined. Sulpizio and Colombo (2017) recently used EEG to examine stress in a single word reading task (lexical decision) in Italian. They manipulated stress neighbourhood consistency as well as stress dominance. In Italian, the first of these is calculated by taking the orthographic ending of a word and dividing the number of words that share the same orthographic ending and have the same stress pattern by the total number of words that share the same orthographic ending. The second simply refers to the most frequent stress pattern. They found very early effects of these, with an interaction between stress consistency and dominance on the P1-N1 complex (70–150 ms) and a relatively similar result at a later time interval (250–350 ms). In both cases, consistency only affected words with dominant stress. In German, Kriukova and Mani (2016) also examined word stress in reading. They did this by first creating a metrical context using either trochaic or iambic words and then presenting a word that did or did not match the context. They found that context only affected iambic words. The differences they found also occurred much later than Sulpizio and Colombo, with significant differences found on the N325 (a marker of metrical processes in speech), N400, and late positive components. They suggested that the results were due to German having more trochaic than iambic words, and that trochaic words have a privileged processing status that reduces contextual effects.

In terms of other reading research, the most similar effect to spelling-stress consistency is spelling-segmental phonology regularity. Sereno, Raynor, and Posner (1998) were the first to examine this using EEG and found no overall effect of it in a lexical decision task. However, when they reanalysed their data based on a subset of participants who showed a spelling-segmental phonology regularity effect in a behavioural task, they found a P200 effect. Fischer-Baum, Dickson, and Fedor (2014) also examined spelling-segmental phonology regularity but only found a significant late positive complex (LPC) effect and even then only in a delayed naming and not a proper name decision task. They suggested that this was due to the correction of segmental phonology that was incorrectly generated. Finally, Yum, Law, Su, Lau, and Mo (2014) examined a similar effect with Chinese characters. They found P200, N400 and LPC effects, but only in a delayed naming and not a character decision task.

Apart from stress in reading, a number of insights can be gleaned from studies examining related areas. In studies examining speech where people listen to words that have been incorrectly stressed, N400 effects have usually been found (e.g. Domahs, Kehrein, Knaus, Wiese, & Schlesewsky, 2009; Molczanow, Domahs, Knaus, & Weise, 2013), suggesting that incorrect stress affects the retrieval of meaning. When illegal rather than incorrect stress patterns are used, they elicit mismatch negativity (Honbolygó & Csépe, 2013). Finally, stress with stimuli that are not incorrect or aberrant in some way have also been examined. Schiller (2006) did this using a stress monitoring task where participants decided in which syllable primary stress occurred with disyllabic stimuli in a go no-go paradigm. He argued that this was more akin to a language production paradigm than other types of experiments using speech perception. He found relatively late occurring N200 effects, with anterior negativity peaking around 475 ms when the first syllable was stressed and 575 ms when the second was. He argued that this was due to the incremental way in which stress was processed, where the stress for words was generated in a left-to-right
manner across syllables (see also Sulpizio, Spinelli, et al., 2015).

If reading in part involves the generation of stress and using it to help extract meaning, there are number of likely overlaps with speech data. With the sublexical route, imputed stress may be used to help access word form and meaning, and may thus affect the N400. In terms of generating stress, given that incremental effects are found in speech, incremental effects may also occur in reading. There may also be effects closer to perceptual ones found due to stress being correlated with pitch in English (e.g. Bolinger, 1958). In this case, if stress changes in the online processing of words in reading affects processes to do with sound processing, then this could affect pitch processing. Such changes could come from sublexically generated stress being initially in conflict with lexical stress that then gets corrected. It could also occur because sublexical stress generation is not always perfect due to conflicting cues within words, and thus for some stimuli, multiple stress patterns are generated which then compete with each other. Since other types of phonological changes that are correlated with stress (e.g. segmental phonology) can cause mismatch negativity (e.g. Näätänen, 2001) and actual changes in pitch also cause anterior effects (e.g. Colombo, Deguchi, Boureux, Sarlo, & Besson, 2011; Schön, Magne, & Besson, 2004; Wickens & Perry, 2015), stress changes in reading might also cause similar results.

This study

Spelling-stress consistency in English was examined in two experiments. Unlike the majority of psycholinguistic experiments in English which have examined disyllables, trisyllabic were used as they have a number of properties that are interesting that cannot be examined in disyllables. A small number of tetrasyllabic stimuli were also used in the second experiment.

An animal decision task was used in both experiments, where participants only responded if they saw an animal word. This was done to examine the extent to which stress is important even in paradigms where participants are not forced to read aloud. Animal decision differs to previous experiments using tasks that do not require speech output which have used lexical or character decision (Sereno et al., 1998; Yum et al., 2014), where people decide whether a string of letters or the character equivalent is a real word. It is, however, more similar to the proper name task of Fischer-Baum et al. (2014) and Taft and van Graan (1998). Although those tasks did not find effects of spelling-segmental phonology regularity, it is worth noting that spelling-stress consistency differs from spelling-segmental phonology regularity. In particular, with spelling-segmental phonology regularity, effects are generally not found in tasks that do not require overt articulation. Alternatively, with spelling-stress consistency, effects are typically found even in tasks that do not require articulation, as shown in the grammatical and lexical decision tasks noted above. It is thus reasonable to assume that spelling-stress consistency effects may exist that can potentially be found.

In the first experiment, a simple manipulation of stress position was done, where words with stress on their initial syllable were compared to those with stress on their second syllable. The basic idea was to see whether a simple bias towards initial syllable stress could be found. There are a number of reasons why an advantage for words with initial syllable stress would be expected.

First, this is what the empirical evidence suggests. In particular, Yap and Balota (2009) report results from a large-scale study that had the reaction times and error rates from thousands of words. After very carefully controlling for many covariates, they noted that in both their naming and lexical decision tasks, words with initial syllable stress had a lower error rate than those with stress on other syllables. This shows that the initial syllable is privileged, although the effect was only found on error rates and thus may be attenuated by factors not directly to do with the processing of stress, as discussed above. Chateau and Jared (2003) found a similar pattern, although it was more limited as only a naming task with 1000 disyllabic words was used. With nonwords, the results are even more convincing as the type of responses can be examined, rather than just reaction times. When Ernestus and Neijt (2008) examined this in a reading aloud task, they found that with trisyllabic nonwords, the majority of responses were given initial syllable stress, and this was true of nonwords with both a “heavy” (48.3% vs. 22.3%) and a “light” (45.5% vs. 28.3%) final syllable structure.

Second, CDP++ (Perry et al., 2013) and the model of Seva et al. (2009) make the prediction that across a large sample of words, there should be a processing bias towards words with stress on their initial syllable compared to those with second syllable stress. In both of those studies, it was suggested that this bias is found due to the greater number of words with initial versus second syllable stress that the models were trained on. Rastle and Coltheart (1998) make a similar argument without the simulations, although they failed to find such an effect in their experiment. Given the relatively weak effects found in other studies and the potential problems measuring stress effects noted above, the
fact that some experiments with comparatively small numbers of items have not found a significant effect is to be expected. The situation here is more complicated, however, because trisyllabic words were examined. According to statistics calculated from the CELEX database (Baayen, Piepenbrock, & van Rijn, 1993), whilst trisyllables still most commonly have stress on their initial syllable, the ratio between initial and second syllable stress is less than with disyllables (Monaghan, Arciuli, & Seva, 2016). Thus, the size of the effect may potentially be reduced compared to studies examining disyllables and thus more difficult to find.

A third possibility is that people use a default stress with some stimuli. That is, some stimuli are likely to be “unmarked” and thus have no strong information that would help determine their stress, and for these people may default to a particular stress pattern such as initial stress (e.g. Levelt, Roelofs, & Meyer, 1999). This could potentially occur within category, such as with stimuli that are known to be nouns. This idea is somewhat similar to the previous one where stress is assigned based on distributional evidence. However, a default form could also be conceptualised as an all-or-nothing form that is used after being abstracted from distributional evidence. In this case, it could potentially function like a simple rule. This makes interpreting the predictions difficult, because unlike current computational models of stress assignment where words with different numbers of syllables are not treated as qualitatively distinct, an abstract rule could potentially do this. Thus, if stimuli are affected by a default form, words with initial stress could behave differently to words with second syllable stress, depending on what the default might actually be – i.e. whether it is syllable number specific, whether it is used for only some syntactic categories, or whether it generalises across all words. Whilst this presents a potentially interesting hypothesis, the extent to which rule-based behaviour has been found in the processing of stress in reading tasks has been relatively limited. For example, in studies looking at reading in Italian, which is far much more conducive to rule-based analysis than English, there is good evidence that more than just rules are needed to explain the data (see Sulpizio, Burani, et al., 2015). Alternatively, the evidence that a default is used, even one based on the general stress distribution of words and not rules, is relatively weak (e.g. Colombo, Zevin, & Garcia, 2009; Krämer, 2009).

A fourth line of evidence from linguistic analyses is how prosodic (metrical) feet (e.g. Hayes, 1995; Selkirk, 1980) are assigned to words. Whilst these data and analyses were not necessarily designed to provide a psychological explanation of how stress is assigned to words when reading, they are relevant for determining which syllables in words are likely to have stress. In particular, in Hammond’s (1999) description of the phonology of English, he examined three major constraints that allow English stress to be predicted based on the feet words are broken down into. His ordering of those constraints meant that the first two syllables in a nonword like pataka would be likely to be given a trochaic initial foot, and thus would have stress on the initial but not second syllable. Pater (2000) came to a similar conclusion using the nonword danaca as an example. Thus, if people use constraints or at least processes that approximate these when assigning stress, then based on these analyses, initial stress in some cases would be preferred over second syllable stress.

Whilst the evidence that was reviewed points to initial stress being favoured in English, some of it is based on how initial stress was defined – basically, counting syllables from the start of a word. When counting from the start, the syllables in a disyllabic word would be assumed to overlap with the first two syllables in, for example, a trisyllabic word. This means stress is more frequent on the initial syllable of words than the second syllable because it most frequently occurs on the initial syllable across all words. An alternative view is that syllable overlap should be based on counting syllables from the end of a word. If this is the case, then the most common type of stress is penultimate. This is because even though trisyllabic words more commonly have initial compared to second syllable stress, disyllabic words most commonly have penultimate (initial syllable) stress, and they are more frequent than trisyllabic words. Thus, penultimate stress is more common than either ultimate (last) or antepenultimate (two to the left of the final syllable) stress.

It is possible to investigate the extent to which counting syllables from the start or the end of a word predicts stress the most accurately. In this case, one prediction is that with trisyllabic words, based on counting from the end of a word, stress will most commonly occur on the second (penultimate) syllable. Based on counting from the start, stress should most commonly occur on the initial syllable. A second prediction based on counting from the end is that words of different syllable lengths should have stress given to them in the same end positions a similar number of times. Neither of these predictions is likely to be correct. The first is in conflict with the linguistic analyses of prosodic feet described above. It predicts nonwords like pataka and danaca should be given stress on the second syllable. If the theories noted were updated so this was true, it would cause problems in their ability to explain large amounts of other data. In addition, as noted above,
Ernestus and Neijt (2008) found trisyllabic nonwords were not given penultimate stress the majority of the time as would be predicted if people counted from the end. They also examined tetrasyllabic nonwords, and found that the distribution of stress based on counting from the end differed to trisyllabic nonwords, which also goes against the prediction that stress is based on counting from the end of a word. When summarising their results, they conclude that English speakers have “preferences for word-initial syllables with primary or secondary stress” (p. 534).

Finally, it is worthwhile noting that none of the evidence presented here suggests that there are not multiple cues that affect the placement of stress in words. Clearly, there are many reasons stress may fall on different syllables, and when investigated, probabilistic effects are found (e.g. Guion, Clark, Harada, & Wayland, 2003). For example, syllable weight and orthographic cues (e.g. Arciuli et al., 2010) are well known factors that affect stress placement. Cues to stress found in one syllable may also affect where stress is placed on different syllables, including those syllables that come before them (e.g. Fudge, 1984). Despite the myriad of effects that affect stress placement, taking the results together, it still seems reasonable to conclude that there is a preference for initial over second syllable stress in English. Whether this can be found in paradigms not reliant on either the articulation of the words and the extent to which the effects can be found in more nuanced ways than simple error rates remains an important question. This was examined in the first experiment using a large and relatively varied group of stimuli.

In the second experiment, spelling-stress consistency based on specific spelling-sound correspondences was examined. This was designed to test the extent to which individual cues are important when reading whilst also tightly controlling for properties of those cues, including those that affect stress across different syllables. This was done by choosing words with the same root morpheme but a different word form. For example, word stress in carbonic is predictable from the final suffix -ic, but it is in conflict with the carbon beginning of the word because most words that begin with carbon like carbonless have initial stress. Because carbonic can be compared with carbonless, a very strictly controlled comparison can be made in terms of spelling-stress consistency because it is only the final letters and their effect on stress that differs. Hence, other types of stress consistency that might occur in the first two syllables are controlled.

There are two potentially non-orthogonal ways that spelling-stress consistency effects could be found with this type of word. One is that if stress is generated incrementally, then a potentially correct stress pattern can be inferred based on relatively high frequency information from the suffix, but only when full information becomes available. For example, if stress was only generated when enough information for each syllable had arrived, carbonic would be likely to first generate car (initial syllable stress), this would be followed by carbon (initial syllable stress since most words with carbon as a root morpheme have that stress), and finally it could potentially be revised to its correct form carbonic if the -ic is used as a cue and not the information from carbon. This could lead to a spelling-stress consistency effect. In this case, incorrect information would have been generated before the correct information became available with words like carbonic and this would thus need to be corrected (see Rastle & Coltheart, 1998, for a similar idea with spelling-segmental phonology regularity).

A second way spelling-stress effects with this type of word could arise is due to inconsistency between the spelling-stress mapping, but this is dependent on the properties of that mapping. In particular, stress is generated using non-linear relationships with the parallel backprop-style model of Ševa et al. (2009) but with linear relationships with CDP++. This means that the model of Ševa et al. would predict that the difficulty in the processing of trisyllabic words like carbonic that have a different stress pattern compared to the root morpheme within them (carbon) should be only slightly higher than words where the root morpheme has the same stress (e.g. carbonless). This is because very high frequency information that is consistent (i.e. the -ic in carbonic) should form an extremely reliable cue as to where stress falls and affect processing more than the much lower frequency root morpheme (see e.g. Plaut, McClelland, Seidenberg, & Patterson, 1996, for a discussion of the role of frequency information in learning in PDP networks). Alternatively, CDP++ only learns linear relationships between spelling and stress and thus the ability for it to take advantage of singular but very reliable cues is weaker as it must additively combine cues from different sources. It is thus more susceptible to conflict from statistical information derived from relatively small groups of items with the same orthographic but differing phonological characteristics. It therefore predicts that even when a very reliable cue exists, there may still be effects of spelling-stress consistency caused by other conflicting spelling-stress relationships. Since neither of these models currently exist in a trisyllabic form, these predictions, whilst interesting, must be taken as speculative.
Experiment 1

Materials and methods

Participants
Fifteen native speakers of Australian English aged between 20 and 30 participated (8 males and 7 females). Participants were initially given information about the study and asked to read and sign an informed consent form that was approved by the Swinburne University Human Ethics Committee. The information included standard exclusion criteria.

Stimuli
The critical stimuli consisted of 360 words that could be evenly split into two groups based on whether they had stress on their initial or second syllable. All had 3 syllables. The two groups were balanced on word frequency, letter length, orthographic neighbourhood (measured with Levenshtein distance, see Yarkoni, Balota, & Yap, 2008), mean bigram frequency, number of morphemes, and concreteness (see Table 1). Concreteness was used as it can affect the N400 (e.g. Kounios & Holcomb, 1994). All measures were taken from the English Lexicon Project (Balota et al., 2007) apart from frequency, stress position, and concreteness. Frequency and stress position were taken from CELEX since the dialect of English spoken by the participants was closer to Received Pronunciation than that used in the English Lexicon Project. Concreteness was taken from Brysbaert, Warriner, and Kuperman (2014), and when matches could not be found but there were very similar words (e.g. single vs. plural form), the alternative form was used. Words were also checked by hand by the author and a collaborator who spoke the same dialect of English to make sure stress in the dialect of the participants was the same stress as given in the database. A further 8 practice words were used at the start of the task and 40 trisyllabic animal names were used as the catch words.

Stimuli presentation
E-Prime software version 2.0 (Psychology Software Tools, Inc., Pittsburgh, PA) was used to present the stimuli. The stimuli were presented in a pseudorandom order in a room with dim lighting. Stimulus triggers were locked to the presentation of each word. In terms of the order of events, a crosshair (+) was first presented for 500 ms. Following this the word was presented. If the word was an animal word, it remained on the screen until a button was pressed, otherwise it was removed after 700 ms. The screen was then blanked for 1000 ms before the next trial began.

Method
Participants were informed about the sensitivity of the EEG to artefacts including eye movement and muscle activity and were asked to remain as still and as relaxed as possible, fixate on the fixation point, and to refrain from excessive blinking, eye, and muscular movements.

EEG data were obtained with a NeuroScan System SynAmps RT amplifier on a Dell Optiplex 780 computer system. The EEG was located in an electrically shielded room. The EEG was recorded with 64 Ag–AgCl electrodes mounted in an elastic cap (International 10/10 System, NeuroCap) sampling at 500 Hz. EEG data was referenced online to the FCz electrode while AFz served as the ground. The EEG signal was recorded online with NeuroScan 4.3 software. The impedance for all the electrode sites was kept below 10 kΩ, except for the CB1 and CB2 electrodes, which were removed entirely.

The data was pre-processed and analysed offline using Fieldtrip (Oostenveld, Fries, Maris, & Schoffelen, 2011) in MATLAB (R2016a). A low pass filter was first applied to the data at 35 hz and bad channels were removed and recalculated based on neighbouring channels with a neighbourhood distance of 2. A time window between −300 and 1000 ms was then used to extract epochs for the trials. From these, the data was decomposed by Independent Component Analysis (ICA, runica algorithm). Subsequently, components identified by the ICA were studied to remove eye movement artifacts, blinks, cardiac rhythm, impedance, and any notable muscular or movement artefacts from the signal. Following this epochs between −100 and 800 ms were extracted from the initial epochs. These were baseline corrected between −100 and 0 and re-referenced to a common average reference. Trials were then removed using the visual artefact toolkit using the following parameters: MaxAbs: 80; Range: 100; kurtosis: 8; zvalue = 1; max z value: 8. The variance parameter was chosen by hand based on visual inspection of the

Table 1. Lexico-statistics for words used in Experiment 1.

<table>
<thead>
<tr>
<th>Group</th>
<th>Log frequency</th>
<th>Letter length</th>
<th>Orthographic neighbours</th>
<th>Mean bigram frequency</th>
<th>Number of morphemes</th>
<th>Concreteness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial stress</td>
<td>1.37 (.72)</td>
<td>8.83 (1.07)</td>
<td>3.23 (.52)</td>
<td>4185 (1169)</td>
<td>2.27 (.62)</td>
<td>2.75 (.93)</td>
</tr>
<tr>
<td>Second syllable stress</td>
<td>1.43 (.86)</td>
<td>8.98 (1.24)</td>
<td>3.11 (.54)</td>
<td>4480 (1282)</td>
<td>2.33 (.74)</td>
<td>2.64 (.86)</td>
</tr>
</tbody>
</table>

Note: Numbers in brackets are standard deviations.
data. Occasionally, when there were obvious outliers these were also removed and a small number of items above these values were left in when they appeared to form part of the main distribution. Statistical analysis was done with the R statistical package using the Afex library. ANOVAs used repeated measures with type III sum of squares and when sphericity constraints were violated, a Greenhouse-Geisser correction was used on the degrees of freedom and $p$ values.

**Results**

One participant who had more than 30% of their data rejected was removed entirely from the analysis. Across the conditions, an average of 18.2% and 18.9% of the data for the words with initial and second syllable stress, respectively, was rejected due to artefacts. The difference was not significant ($t < 1$).

The data was first initially visually inspected to help determine time windows with which to compare mean results across the conditions. Effects were chosen based on this inspection as well as with respect to the previously reviewed literature. To examine the P200, a window was created by first finding the time point between 150 and 250 ms at which the highest $\mu$V value occurred in the anterior electrodes and the minimum $\mu$V value occurred in the posterior electrodes (see Figure 2 for the actual electrodes used). After this, the average of the time points was found (198 ms) and results 40 ms to each side of it were used to create the window. This window clearly incorporated the P200 across all of the conditions examined. Further inspection of the data showed a divergence between conditions around 350–450 ms, and this window was used to examine the N400. Finally, a LPC window of 500–800 ms was used.

In terms of regions to examine, inspection of the topographic maps showed that in the P200 and LPC window, the data showed two main clusters of activation, which were largely split between anterior and posterior electrodes. To examine these, the time-windows were examined using 3 anterior-posterior areas and 3 regions based on laterality (see Figure 2). Data analysis for these effects was done using an initial 3 Region (Anterior/Center/Posterior) $\times$ 3 Laterality (Left/Mid/Right) $\times$ 2 Stress Position (Initial Stress/Second Syllable Stress) ANOVA. In the N400 window, there was a positive going topography in centro-posterior regions. Therefore, using an ANOVA examining different groups of electrodes like the other comparisons was not used, but rather, a group of centro-posterior electrodes that covered this region were chosen. Topographic maps based on the average data within the time windows used, mean results, and example electrodes appear in Figure 3.

**P200**

The results of the ANOVA showed there was a significant Region $\times$ Stress Position interaction ($F(1.24, 16.10) = 6.39, p = .016, \eta^2_p = .11$), with the words with stress on their initial syllable showing a greater amplitude than those with stress on their second syllable. As can be seen from the topographic maps, differences in stress position appeared to be divided between posterior and anterior regions. The data was therefore examined in the different regions separately. The results showed Stress Position was marginally significant in the anterior region ($F(1, 13) = 4.56, p = .052, \eta^2_p = .068$) and was significant in the posterior region ($F(1, 13) = 9.50, p = .0088, \eta^2_p = .18$).

![Figure 2](https://example.com/figure2.png)

*Figure 2.* Regions used to analyse the data in both experiments.
An ANOVA examining the data showed that Stress Position was significant ($F(1, 13) = 5.01, p = .035, \eta^2_p = .089$), with words with initial stress showing a greater N400 amplitude than those with second syllable stress.

**LPC**

There were no significant effects of interest.

**Discussion**

The results show that words with initial stress differed significantly to words with stress on their second syllable in both the P200 and N400 windows. Thus, they demonstrate that significant effects of spelling-stress consistency can be found, even using words chosen based on the simplest definition of stress consistency.

**Experiment 2**

The stimuli used in the first experiment used a consistency measure simply based on stress position. A different type of consistency measure can be used that is also theoretically interesting. In particular, as discussed above in *This study*, there are words that have part of their structure that is inconsistent, but where the final form can be inferred accurately due to a suffix that is strongly predictive of the stress and occurs at a high frequency (e.g. *carbonic*). In this experiment, stress consistency was defined based on words having a two-syllable root morpheme that mainly did (consistent; e.g. *carbon-*, *abduction*) or did not (inconsistent; e.g. *carbonic*, *abductee*) use the same stress as other words starting with the same morpheme. Like the first experiment, words that had stress on either the initial or second syllable were used. However, based on the first experiment where words that start with initial stress appear to be somewhat privileged, a number of 4 syllable words were used where the final two syllables formed a suffix that was stress predictive (see Fudge, 1984). This was necessary due to the number of stimuli available that could be used without repeating the same patterns excessively and given the relatively small size of the effect in the previous experiment.

**Materials and method**

**Method**

The method was the same as the previous experiment.
Participants
Fifteen people between the age of 20–30 participated in the study (8 males; 7 females). All were native Australian English speakers.

Stimuli
The critical stimuli consisted of 192 words that could be evenly split into a 2 Stress Consistency × 2 Stress Position design. Stress Position was whether the words had stress on either their initial or second syllable. The words were balanced on log frequency, letter length, number of syllables, orthographic neighbours measured with Levenshtein distance scores (Yarkoni et al., 2008), number of morphemes, bigram frequency, log sum frequency of words that start with the root morpheme, number of words which shared the same root morpheme excluding words that had a frequency of zero, log frequency of the orthographic ending, concreteness, and proportion of derivational/inflectional forms. Three measures of spelling-stress consistency were calculated by dividing counts derived from words that shared the same root morpheme with the same stress pattern by those that shared the same root morpheme with any stress pattern. The measures were based on simple word counts excluding those with a frequency of zero, summed frequency of the words, and log summed frequency of the words. With the simple word counts, words of zero frequency were eliminated as they were often relatively obscure and could potentially bias simple counts. Root morpheme frequency was calculated by matching the root morpheme with all other words that started with the morpheme that were less than 5 syllables long. The syllable limit was used for convenience to remove very low frequency forms (including idioms and compound words divided by a space) in the database. When words contained final letters (e.g. “e”) that were not used in many morphologically complex forms, matches were allowed without the letter (e.g. combine and combination). Orthographic ending frequency was calculated by matching the end letters from all trisyllabic words in the database. All scores were generated from the CELEX database except orthographic neighbourhood and bigram frequency which were taken from the English Lexicon project (Balota et al., 2007) and Concreteness which was taken from Brysbaert et al. (2014). A summary of the statistics appears in Table 2.

Results
The P200 window was calculated in the same way as the previous experiment, with the peak occurring at 206 ms.
Further inspection of the data suggested that it was reasonable to keep the same N400 (350–450 ms) and LPC (500–800 ms) time windows as the previous experiment. Across the conditions, an average of 16.8%, 16.4%, 15.4%, and 16.8% of the data for the Consistent/Initial Stress, Consistent/Second Stress, Inconsistent/Initial Stress, and Inconsistent/Second Stress groups, respectively, was rejected due to artefacts. The differences were not significant ($F < 1$). The results appear in Figure 4.

**P200**
There were no significant results of interest.

**N400**
A 2 Stress Position (Initial/Second Stress) × 2 Stress Consistency (Consistent/Inconsistent) ANOVA examining the data showed a significant Stress Consistency × Stress Position interaction ($F(1, 14) = 11.01, p = .005, \eta^2_p = .19$). Inspection of the data showed that appeared to be because there was a morphemic effect, where words with a root morpheme with initial stress (e.g. *carbon*) showed an N400 effect with a greater amplitude than those with second-syllable stress (e.g. *abduct*), regardless of the whole-word phonology.

**LPC**
There were no significant results of interest.

**Discussion**
The results differed compared to the previous experiment, with no significant P200 effect being found. Alternatively, there was still a significant N400 effect. However, it was not the same pattern as the first experiment where words with initial stress showed a greater amplitude than those with second syllable stress. In this experiment, words with a root morpheme that typically has initial stress (e.g. *carbon*) appeared to show an N400 with a larger amplitude than those words with a root morpheme that typically has second syllable stress (e.g. *abduct*), regardless of the stress pattern of the final word form.

**General discussion**
The idea that phonology is important in reading is well established, but the role which word stress plays is not so clear, with mixed results being reported. Here, two experiments were run examining the extent to which the position of stress in a word and the consistency of...
spelling-stress relationships affects reading. This was done in a task that does not require speech output, and thus does not force stress to be processed. Effects of both stress position and spelling-stress consistency were found, although in the second experiment, the effect occurred based on morphemes within words, rather than the whole-word stress pattern. The results are important as they show that effects of word stress can be found in English relatively early in processing (P200) and in the processing of meaning (N400), although no significant LPC effects were found. The results in the time windows examined allow a number of insights into how stress is generated from orthography.

**P200**

In the first experiment, the results showed that there was an effect of stress position in the P200 window. Differences in the P200 can come from many different aspects of language processing, such as the context words are presented in (e.g. Dambacher, Kliegl, Hofmann, & Jacobs, 2006), ease of phonetic processing (e.g. Strauß, Kotz, & Obleser, 2013), and spelling-segmental phonology inconsistency (Sereno et al., 1998; Yum et al., 2014). The results here are consistent with Sereno et al. in that the stress-inconsistent words showed a lower amplitude than the consistent ones (Yum et al. found a more complicated pattern). The P200 could have been elicited at multiple levels of processing. For example, it might be due to differences in the initial output of the sublexical system or that output affecting other levels such as phonetic processing, where lower quality stimuli produce a P200 with a lower amplitude (e.g. Strauß et al., 2013).

Based on an analogy with how CDP++ works, there are at least two possible ways the P200 could be affected by stress before speech output. One is that the P200 is caused by differences in the speed and level at which the stress nodes are activated, with the initial syllable stress node rising more quickly than the second syllable stress node. Alternatively, it could be caused because words with second syllable stress may cause more activation across both initial and second syllable stress nodes. This occurs because of the way CDP++ learns, where the error of words presented to the model does not necessarily ever get minimised to zero. This means that words that are more difficult to learn, which tend to be those with second syllable stress due to their lower frequency of occurrence, cause activation over more than just the stress node they are associated with. Thus, whilst the stress node they are associated with may have a lower activation than with words that are easy to learn, the total activation of both nodes may be higher. The current results cannot distinguish between these two possibilities. The model of Ševa et al. (2009) is more constrained in this respect because it only uses one output node for stress, and thus cannot predict that the summed activation of nodes may cause a difference. However, it would be simple to change that model to have one output node for each syllable, like CDP++, should further data support this possibility.

Actually why the output of the sublexical system shows a difference between initial and second syllable stress in this experiment is open to a number of interpretations. One possibility is that without explicitly controlling for many characteristics of the words, the sublexical stress mechanism tends to give pronunciations with initial stress due to such words being more common and thus having more orthographic cues that help predict initial stress. Thus, it may represent a system that is simply biased towards initial syllable stress due to the frequency at which words with different types of stress occur, as predicted by CDP++ and the model of Ševa et al. (2009).

To further investigate whether the stimuli used in Experiment 1 had potential orthographic cues that could have biased processing towards initial or second syllable stress, the stimuli were post-hoc examined on a number of different cues. These cues were taken from those described in Monaghan et al. (2016). They were also calculated the same way, using token frequency counts with phonological and orthographic forms from the CELEX database and with separate calculations done for words with different numbers of syllables. In terms of measures, the proportion of times the letters in the body of the initial and last syllable of the words was associated with either initial or second syllable stress was calculated. Similarly, the extent to which letters from the start and end of the words are associated with initial and second syllable stress was also examined (from one to five letters). Difference scores between the proportions derived for each of the initial and second syllable word stress groups were then calculated. This gives a measure of the proportion of times a particular word stress is predicted from the orthographic cues of a word. The difference between the words with initial syllable stress was then added to the difference between words with second syllable stress. This gives a measure of the overall bias towards initial or second syllable stress across the two word groups. The results from this appear in Table A1. As can be seen, across a number of the metrics examined, there was a bias towards cues favouring initial syllable stress. This supports the possibility that orthographic cues in the
stimuli set favoured words with initial over second syllable stress, as the overall pattern of statistics counted from all words does too.

The results from the second experiment where significant differences in the P200 were not found are also informative. In that experiment, the initial letters of words co-occurred in clusters of words with the same initial spelling. Thus, they may have acted as small islands of regularity due to the identical sets of correspondences having a comparatively high frequency of usage compared to more randomly selected words. Such islands can be relatively robust to the effects of more general statistical patterns that exist across larger sets of stimuli (see e.g. Albright, 2002), and hence the overall bias towards the processing of words with initial stress may have been eliminated.

To further examine this possibility, potential orthographic cues were examined in the stimulus set of Experiment 2. The results appear in Table A2.1 As can be seen, the words displayed the expected pattern where consistent and inconsistent words were biased towards the type of stress that would be expected based on those categories. In terms of overall bias, however, unlike the first experiment, the cues were generally more predictive of second syllable than initial syllable stress. The differences favouring second syllable stress in this experiment were weaker than those favouring initial syllable stress in the first experiment, however. These results suggest that the words chosen were unlikely to be representative of the database as a whole or like the first experiment where the cues were biased towards initial syllable stress. This supports the possibility that the P200 was driven by orthographic cues because differences were only found in the first experiment where the orthographic cues were the most predictive of one type of stress over the other.

An alternative interpretation to words with initial stress being special due to their frequency is that stress is generated as an incremental process. If this occurred syllable-by-syllable, for example, there would always be an initial bias towards words with initial stress because single syllables out of context are always stressed, and they may thus need to be corrected if further conflicting information was processed (see This study above). However, since no P200 difference between words with initial and second syllable stress was found in the second experiment, it would suggest that at least two syllables of information were processed together (i.e. the root morpheme) before additional incremental information became available.

A third possibility is that the origin of the effect is at the level of speech output, and not the generation of spelling-stress information in the reading system, and this can affect processing even in silent reading tasks. For example, people may prepare a metrical frame (e.g. Roeloffs & Meyer, 1998) early in processing that uses initial stress because words with initial stress are the most common. Differences could then arise if this frame is revised because activation in the stress output nodes differs or if activation can feed back from the metrical frame to the stress output nodes. However, based on the results of the second experiment where no P200 differences were found, it is possible to rule this out. In this case, if there was just a simple default to initial stress for metrical frames, such a difference between the experiments should not have been found.

N400

In both experiments, an effect of stress position was found on the N400. This is interesting because the P200 effect occurred only in Experiment 1. Thus, an early processing difference was not necessary for later semantic processing differences to be found. This suggests that it is likely that there are processing advantages of having initial stress that affect different levels of processing, and they are not simply based on early lexical access differences that then carry on to the later processing of meaning.

The results of the second experiment offer particular insight into the processing dynamics that may be occurring. In particular, it was not the final word form which was important in helping words access their meaning faster, but it was the stress of the initial disyllabic root morpheme. In this case, words that used a root morpheme that generally had initial stress showed a N400 effect with a greater amplitude than those that did not. Thus, the constituents that words can break down into was important.

One possible explanation of this is that it is a task specific effect, where participants simply read the disyllabic morpheme at the start of the words, and then based their responses on this. In particular, they may have assumed that if the initial two syllables of a stimuli were not a word, then the stimuli must have been an animal. Thus, for example, when presented with “carbolic”, carbon would be read and this is a word and hence not an animal, and when reading elephant, ele would be read and this is not a word and thus must be an animal. Whilst this might be possible, there are a number of potential problems with it.

First, with English spelling, the morphemes within tri-syllabic words are not always identical to the disyllabic morphemes within them. This commonly happens with the letter –e and –y and there are also many idiosyncratic cases (e.g. install-instalment, decent-decency). For
example, it was assumed here that *oppose*, *opposable*, and *opposed*, all share the same morpheme. However, *opposable* misses the final –e of *oppose*. In the non-animal stimuli used here, the bound morphemes were not identical to the unbound ones 40% of the time. Thus, to use this strategy, it would need to be assumed that there can be different orthographic forms of bound morphemes as well as the unbound form. For example, people would need to have separate orthographic representations of *oppose* (unbound/bound) and *oppos* (bound) to allow the processing of *opposed* and *opposable* this way. Note that there are models which clearly allow for partial activation of meaning from parts of morphemes that do not have this problem (e.g. Milin et al., 2017). However, these models do not use any early morpheme-like orthographic segmentation. They thus have no obvious method for allowing people to quickly recognise words based on internal constituents and hence use this as a strategy in the task.

A second problem with assuming people can simply read off the initial bound/unbound morphemes without being affected by the suffix is that it assumes people can strategically ignore parts of words when reading. This seems unlikely based on a number of sources of evidence. First, reading words that do not have spaces between them slows reading considerably (e.g. Epelboim, Booth, Ashkenazy, Taleghani, & Steinman, 1997). Thus, if just normal reading between words is affected by spacing, it seems reasonable to assume that trying to entirely ignore parts of words within words that are not necessarily simple to identify is also likely to be difficult. Second, current models of morphological processing (see e.g. Taft, 2015, for a review) that assume words can be broken down into morpheme or morpheme-like units also assume they are processed automatically. If orthography was not processed automatically, it would be difficult to explain the very large amount of data using fast masked priming examining such decomposition (e.g. Rastle & Davis, 2008). Finally, studies directly examining the automaticity of different parts of the word recognition system have also found good evidence that orthographic forms are processed automatically (e.g. Reynolds & Besner, 2006). This again suggests that breaking a word down into parts and then processing those parts whilst ignoring others is likely to be more difficult than simply processing the words like normal and making a decision based on the whole-word form.

Given the different reasons people may have problems just focusing on the initial morpheme, the possibility that this is responsible for the data pattern seems unlikely. However, if people really can simply modify their strategy to use the initial morpheme, then the results found would suggest that there is a bias towards initial syllable stress on these morphemes. Some support for this possibility comes from an analysis of the orthographic cues of the morphemes (see Table A3), which tend to favour initial stress.

An alternative explanation of the results is that word stress affects which neighbours of words are activated – that is, when words are read, they not only activate their own lexical form, but other words that are visually and phonologically similar to them (e.g. Coltheart, 1978; Yarkoni et al., 2008). The results here suggest that the activation of these neighbours may in part be affected by word stress. In this case, due to an initial syllable stress bias, from, for example, orthographic cues that are present in the initial disyllabic morphemes (see Table A3), this could have caused more neighbours to be activated with initial compared to second syllable stress. This would explain the pattern of results found because words with many neighbours cause a larger N400 than those with few (e.g. Holcomb, Grainger, & O’Rourke, 2002; Laszlo & Federmeier, 2011). Since the initial stress bias was only present in the disyllabic morphemes within the words and the onsets of the words, it would need to be assumed, as both CDP++ and the model of Ševa et al. (2009) do, that overlap is calculated based on words of any syllable number rather than being syllable number specific. It is worthwhile noting that this explanation is not necessarily in opposition to the explanation proposed for differences in the P200 between Experiment 1 and Experiment 2, where orthographic cues were implicated. This is because it suggests that there is an advantage for initial over second syllable stress, but this may not be found when other factors obscure it, such as when transitory states cause both stress nodes to be activated before one wins out.

How the neighbours are activated is worth considering. One possibility is words are initially split into morpheme or morpheme-like parts (e.g. Rastle & Davis, 2008; Taft, 2001) and these parts are then processed largely separately. If this happened, a word like *carbonic* would activate its individual parts (*carbon* + *ic*). This would cause the root morpheme to be more important than if the word was not split because it would have access to its own representation. However, if this happened, some way of getting the whole word-form phonology back from the parts would be needed for inconsistent cases. Unless the morphemic parts access their semantics and other features and then reconverge on a whole-word orthographic form before phonological processing, this possibility does not seem especially likely. This is because it predicts there should be a large late cost in over-riding the incorrect phonology produced from the constituents, but this was not found.
Alternatively, if letters or groups of letters are processed incrementally, then the results are potentially congruent with an account where morphemic decomposition is not initially needed. In this case, the carbon in carbonic would first be processed with initial syllable stress, and hence may be more likely to activate initial syllable stress neighbours. Only later would the phonological form be corrected. The cost of correcting the inconsistent form would not be high, however, because the incremental processing would allow some level of automatic self-correction once the final -ic was processed. The stress neighbourhood effect would also be made stronger with a subsyllabic mechanism such as that CDP++ uses because in that model cues can only be weighted in a linear fashion. Because of this, with the inconsistent words, the model cannot take full advantage of the single highly predictive high frequency suffix the words finished with because they were accompanied by the letters in the initial morpheme that predict the opposite stress pattern. Thus, even if the stress for the inconsistent words with second syllable stress can be corrected by the suffix, it is likely there would be partial activation still predicting initial stress, hence helping initial stress neighbours remain activated.

Apart from the incremental generation of stress and feedback from sublexically generated stress, there are at least two other ways that stress could affect the way neighbours are activated. One is that neighbours are activated orthographically which in turn activate their corresponding phonological neighbours. These could then contribute to stress at the output level (stress nodes) which would then feedback and activate neighbours with a similar stress. If this is the case, the stress effect may be in part due to lexical activation. A second possibility is that the phonology of words commonly differs depending on how they are stressed (e.g. cf., fatal, fatality), and this could increase the difference between initial and second syllable stress neighbours due to differences in the effect of segmental phonology. This would occur due the left-to-right nature in which segmental phonology is processed (e.g. Perry & Ziegler, 2002; Rastle & Coltheart, 1999; Weekes, 1997) and because initial letters are more commonly associated with initial compared to second syllable stress. This effect could be quite long lived in stimuli sets like that used in Experiment 2. This is because the majority of the disyllabic root morphemes used had orthographic cues that are biased towards initial stress at the beginning of their letter sequences. Thus, the segmental phonology initially generated sublexically would be likely to reflect phonology associated with initial stress. It would only be later in processing that enough of a graphemic context would be generated such that the phonology generated could become more typical of words with second syllable stress.

The explanation of CDP++ can be compared with the backpropagation model of Ševa et al. (2009). In that model, the words are not only presented in parallel but the model can learn nonlinear relationships between spelling and stress. Because of this and because the stimuli used here had suffixes that were of high frequency and were reliable predictors of stress, the amount of error when learning these words would be minimised. It is thus not clear how it could predict the pattern of the data, since the spelling-stress consistency effect should be minimised in the model. Thus, it would presumably predict a null-effect of spelling-stress consistency due to learning, which is not what was found. Even if it could produce a spelling-stress consistency effect, it is not clear how it could predict that the effect should occur with root morphemes rather than with whole words.

**LPC**

Studies examining spelling-segmental phonology regularity have found LPC effects and suggested they were due to the correction of incorrect phonology (Fischer-Baum et al., 2014; Yum et al., 2014). This study found no effect of spelling-stress consistency in this window. One likely reason for this is that the stimuli here were not chosen to be exceptionally difficult (i.e. very inconsistent), and, in the second experiment, the stress patterns were generally predictable based on at least one strong cue. Thus, unlike some studies examining spelling-segmental phonology, this study did not use the most extreme possible manipulation and thus the need for any large phonological correction would have been more limited. It is also the case that strong effects would not necessarily be expected based on results reported in Fischer-Baum et al.’s study on spelling-segmental phonology regularity in English. The strongest effect they found was only significant at $p = .023$, despite a very strong manipulation with their stimuli.

It is possible to glean some insight into the processes that might underlie spelling-sound consistency effects given that effects of spelling-stress consistency were found in the earlier windows here (P200 and N400), but Fischer-Baum et al. (2014) only found spelling-segmental phonology regularity effects in their LPC window. In particular, current models of reading assume a continuous activation mechanism whereby stress and segmental phonology patterns are generated continuously and automatically corrected over time. Because of this automatic correction, even CDP++, which allows stress to be corrected relatively late.
compared to the model of Ševa et al. (2009), could not generate such a late effect (i.e. much later than the time it takes to extract meaning). However, a possible reason for the discrepancy between the models and the data exists that is still consistent with the models. In particular, the results in Fischer-Baum et al. may represent words where the final phonological form is incorrectly generated when it is not corrected via the standard automatic means. In this case, a self-monitoring mechanism (e.g. Levelt et al., 1999) may be used that allows higher-level processes to capture this type of error and then recompute the final form. Such a process would occur very late compared to the automatic ones suggested by models of reading, since it would act upon the final phonological form generated by the models rather than earlier more automatically generated ones.

The lack of LPC effects also offers some insight into why no anterior effects were found, assuming these would have occurred from large pitch changes caused by the correction of incorrectly generated stress patterns caught via self-monitoring. With the data here, the number of times incorrect stress like this was actually generated and corrected may simply have been relatively small as suggested by the lack of significant differences, and the effect may simply not be strong enough to find easily. In terms of the earlier and more automatic processes, these may not generate such large effects compared to when changes from a fully processed output form are made and are thus not evident in the data here.

Further issues

Throughout the discussion, the idea of incremental processing has been used to offer potential explanations for the results. This idea has been extensively examined in spelling-segmental phonology translation (e.g. Perry & Ziegler, 2002; Rastle & Coltheart, 1999; Weekes, 1997). However, spelling-stress translation has number of different characteristics, and it is unclear what units might be used for the incremental generation of stress. For example, sublexical stress could appear as soon as any of the letters in a word are processed. However, even within syllables, this would lead to large amounts of incorrectly generated stress activation, because an important predictor of whether English syllables are stressed are the final coda consonants (e.g. Hayes, 1982) and these would only become available after the initial letters to do with the onset and the vowel had been processed. Thus, generating stress before full syllabic information is available can be misleading. Similarly, it is well known that adjacent syllables can affect the stress which is assigned to them (e.g. Fudge, 1984), something which was taken advantage of in the second experiment here. Given this, it may be worthwhile holding back the generation of stress until more than even a single syllable of information becomes available. This would still allow incremental effects to emerge, but the time-course for stress generation would be different. Whether the beginning of incremental generation in reading using larger units is based on syllable information or larger units that are rapidly obtained from early orthographic information (e.g. Rastle & Davis, 2008; Taft, 2001) cannot be determined from this study, although there was some evidence that at least two syllables of information in root morphemes is processed before further stress information becomes available.

Cross-language differences

The results of this study differ in important ways from those reported in Italian by Sulpiizio and Colombo (2017). Most notably, they found differences on the N1-P1 complex, which is a very early effect. Since it is hard to find effects on the N1-P1 complex when only small numbers of stimuli are used (e.g. Woodman, 2010), this offers some insight into how strong the effect is in Italian compared to English. In particular, Sulpiizio and Colombo used 60 items for each main effect and 30 for individual comparisons and found a significant difference between groups. Here, 180 stimuli were used in each group in the first experiment and there was no significant effect in the data in this window (unreported statistical comparisons that were made confirm this). This suggests that early stress effects are likely to be very robust in Italian, unlike English, and that the time-course of stress effects may well differ in the two languages. This offers an explanation for cross language differences discussed in the introduction, where it was noted that spelling-stress consistency effects in Italian are generally found in reaction times, unlike English. In this case, if there is a response criterion that people use before responding to words and it is generally set within limited bounds, the chance that stress effects in Italian would have been resolved within these bounds compared to English would be higher due to their earlier emergence. This would allow effects in reaction times to occur in Italian, whereas in English, the resolution of stress consistency effects may come outside the response bounds and hence tend to cause effects on error rates.

Compared to the Italian results, the results here are more similar to the German ones of Kriukova and Mani (2016). In both their study and the two experiments
reported here, an N400 effect was found. Alternatively, the results from Experiment 1 here showed a P200 effect which they did not report, and their results displayed an N325 effect which was not found here. There were differences between the studies that may have contributed to these differences. Most notably, the paradigm they used created a metrical context before each target word which would have increased the effect of metrical conflict compared to this study. This would explain the N325 they found, which they noted is associated with conflict in metrical stress. In terms of the P200, it is not clear whether their stimuli would have been biased in the same way as the stimuli in Experiment 1 here, where differences in the predictability of stress caused by orthographic cues were found. Despite these differences, both their study and this study suggest that stress is processed differently depending on whether it falls on the initial or second syllable. These results concur with the behavioural work examining both German and English stress done by Ernestus and Neijt (2008) who proposed that people have a preference for initial syllable stress in both of these languages.

**Conclusion**

This study examined the extent to which the reading of words is affected by inconsistent spelling-stress relationships and stress position. Significant effects were found on two different components (P200, N400) but no LPC effect was found. These results are interesting because they show that the computation of word stress is important when reading, even in a task that does not require overt articulation and in a language that has complex spelling-stress relationships. The results are also interesting because they suggest that some type of incremental processing might occur in the assignment of stress and they thus help elucidate possible computational processes that are used.

**Note**

1. Note that initial letter scores are not exactly the same across groups using the same root morpheme because scores for words are calculated based on only those words with an identical number of syllables. Thus, the same set of letters can have different scores if it occurs in words with a different number of syllables. There are also a small number of items that are different due to idiosyncratic CELEX codings that can, for example, cause some words to have different orthographic and phonological syllables despite what appears to be the same orthography (e.g. the “neg” in negation and negative).

**Acknowledgement**

I would like to thank an anonymous reviewer for helpful comments.

**Disclosure statement**

No potential conflict of interest was reported by the author.

**Funding**

This work was in part funded by the Australian Research Council from a Discovery grant awarded to Conrad Perry [DP170101857].

**References**


Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). Psychophysiology, 38, 1–21. doi:10.1111/1469-8986.3810001


