Article

Language and Speech

Language *and* Speech
56(1) 23–44

DOI: 10.1177/0023830911434277
las.sagepub.com

Perception and Bias in the Processing of Compound versus Phrasal Stress: Evidence from Event-related Brain Potentials

Stewart M McCauley
Cornell University, USA

Arild Hestvik and Irene Vogel
University of Delaware, USA

Abstract

Previous research using picture/word matching tasks has demonstrated a tendency to incorrectly interpret phrasally stressed strings as compounds. Using event-related potentials, we sought to determine whether this pattern stems from poor perceptual sensitivity to the compound/phrasal stress distinction, or from a post-perceptual bias in behavioral response selection. A secondary aim was to gain insight into the role played by contrastive stress patterns in online sentence comprehension. The behavioral results replicated previous findings of a preference for compounds, but the electrophysiological data suggested a robust sensitivity to both stress patterns. When incongruent with the context, both compound and phrasal stress elicited a sustained left-lateralized negativity. Moreover, incongruent compound stress elicited a centro-parietal negativity (N400), while incongruent phrasal stress elicited a late posterior positivity (P600). We conclude that the previous findings of a preference for compounds are due to response selection bias, and not a lack of perceptual sensitivity. The present results complement previous evidence for the immediate use of meter in semantic processing, as well as evidence for late interactions between prosodic and syntactic information.

Keywords
Compounds, event-related potentials (ERPs), meter, prosody, stress

I Introduction

In previous psycholinguistic work on speech prosody, one phenomenon that has received little attention is the use of contrastive stress patterns to distinguish meanings at the suprasegmental

---

Corresponding author:
Stewart M McCauley, Department of Psychology, Uris Hall, Cornell University, Ithaca, NY 14853-7601, USA.
Email: smm424@cornell.edu

level. Setting non-neutral stress patterns aside (such as those providing emphasis, contrast, and focus), there are two types of information imparted through contrastive stress in English: differences between single words (lexical stress) and differences between compound words and phrases (compound and phrasal stress). The two types of stress contrast can be exemplified by minimal pairs, such as fórbear (noun) vs. forbéar (verb) in the case of lexical stress, and gréenhouse (compound) vs. green hóuse (phrase) in the case of compound/phrasal stress. Previous studies using behavioral methods (e.g., Cutler & Otake, 1999; Cutler & Van Donselaar, 2001; Soto-Faraco, Sebastián-Gallés, & Cutler, 2001) and electrophysiology (e.g., Friedrich, Alter, & Kotz, 2001; Friedrich, Kotz, Friederici, & Alter, 2004a) suggest that listeners use lexical stress information during spoken word identification. However, the distinction between compound and phrasal stress and the role it plays in online comprehension remain relatively unexplored, and represent the focus of the present study.

Whereas lexical stress often involves simultaneous segmental and suprasegmental change (e.g., note the difference in vowel quality for the first vowel in cónvict [noun] vs. convíct [verb]), compound/phrasal stress variation is expressed only in the suprasegmental domain (Cutler, 1986; Vogel & Raimy, 2002). In compounds, the first element tends to be the primary stressed syllable; in phrases, both elements tend to bear primary stress, with the second element being stronger than the first (Gussenhoven, 2004; Plag, Kunter, Lappe, & Braun, 2008). It has also been reported that minimal pairs of phonetically identical compounds and phrases (e.g., bláckboard vs. black bóard) tend to differ in length, with the phrase being slightly longer (Farnetani, Torsello, & Cosi, 1988).

In addition to work on the acoustic features of the compound/phrasal stress distinction, there has been research focusing on its perception and processing, much of which has used minimal pairs of segmentally identical but prosodically distinct phrases and compounds (such as hot dóg and hótdog). Farnetani et al. (1988) found that, when subjects were asked to identify such items as either compounds or phrases, compounds were rarely mistaken for phrases, while phrases were often mistaken for compounds with a high degree of confidence. Vogel and Raimy (2002) used a picture/word matching task in which subjects were presented with pairs of images representing a compound and the corresponding phrase, followed by sentences in which either the compound or the phrase appeared, in a neutral context. Subjects were asked to indicate which image matched the sentence. Children and adults displayed a greater tendency to interpret phrasally stressed items as compounds than the reverse. However, the opposite pattern for adult subjects was observed when novel compounds were used; adults tended to disregard a compound stress pattern when they did not have a lexical compound corresponding to an item they encountered for the first time. Vogel, Hestvik, Bunnell, and Spinu (2009) employed a similar task in a study with a large number of adult subjects, observing the same pattern of greater accuracy for compound stress with both synthetic and natural speech stimuli.

These studies, which have all relied on offline measures of comprehension, raise a number of questions about the contrastive use of compound and phrasal stress which may be easier to address using online measures with high temporal resolutions, such as electrophysiology, which hold the potential to differentiate between perceptual and post-perceptual processes. The present study employed electrophysiology to investigate whether the observed preference for compounds stems from poor perceptual sensitivity to the compound/phrasal stress distinction, or from post-perceptual bias in behavioral response selection (e.g., due to frequency, plausibility, or a preference for analyzing strings as words). A secondary goal was to illuminate the nature of the compound/phrasal stress contrast's contribution to online comprehension.

Before we describe the details of the present study, we first review previous findings from the electrophysiological literature on speech rhythm/meter. Predictions for the current study are then introduced in light of this research.

1.1 Electrophysiological correlates of metrical/rhythmic perception and use in online processing

Previous electrophysiological work has suggested a role for metrical and rhythmic information in a wide range of language processes. Early work in this direction investigated the role of rhythmic information in spoken word identification. Friedrich et al. (2001) found that pitch contours within words modulate early auditory potentials, suggesting that stress information is automatically discriminated within the first syllable during spoken word recognition. Friedrich et al. (2004a) extended these findings, showing that pitch contours within words modulate a positive deflection known as the P350, which has been linked to facilitated lexical identification (see also Friedrich, Kotz, Friederici, & Gunter, 2004b).

Investigating the role of meter in syntactic processing, Schmidt-Kassow and Kotz (2008) found that the duration of constant intervals between successive phrases in a sentence modulated the latency of the P600, a late positivity tied to violations of tense, agreement, and phrase structure (e.g., Hagoort, Brown, & Groothusen, 1993), as well as difficult syntactic integration (Kaan, Harris, Gibson, & Holcomb, 2000). Extending these results, Schmidt-Kassow and Kotz (2009a) demonstrated an anterior negativity in response to metric and combined metric/syntactic violations, which deflected earlier than an anterior negativity elicited by syntactic violations alone and was followed by a late posterior positivity (P600). The authors took this as evidence that metric information is processed early and used as a grid to organize the incoming speech stream, and that metric and syntactic cues interact in a later “integrational” stage.

Other work has focused on the interplay between meter and semantics. Magne, Astésano, Aramaki, Ystad, Kronland-Martinet, and Besson (2007) found that misplaced stress accents in French elicited an N400, a negative component linked to semantic processing (see Kutas & Federmeier, 2000, for a review). The authors interpreted this effect as reflecting disrupted access to word meaning brought about by the changes to words’ metrical structures. Importantly, this effect was present regardless of whether the task was explicit towards semantics or meter, suggesting that metrical information is automatically used in semantic processing. A follow-up study found that musical expertise modulated this component, in addition to enhancing an early exogenous component, the P200 (which reflects perceptual processing; Hillyard & Picton, 1987; Shahin, Roberts, Pantev, Trainor, & Ross, 2005), in response to the same metrical incongruities (Marie, Magne, & Besson, 2011). These findings have also been extended to silent reading. Magne, Gordon, and Midha (2010) found that metrically unexpected words (stressed on the second syllable instead of the first, as expected, or vice-versa) in visually presented lists elicited an N400-like negativity, which the authors interpret as reflecting the impact of the unexpected stress pattern on semantic processing. Luo and Zhou (2010) found that abnormal rhythmic patterns of the verb-noun combination in visually presented Chinese sentences elicited an early positivity, an N400-like negativity, and a late positivity, with all three components modulated by semantic congruency, which the authors take as further evidence that rhythmic patterns are used in semantic integration during silent reading.

In order to explore misplaced stress independently of semantic processing, Rothermich, Schmidt-Kassow, Schwartze, and Kotz (2010) exposed subjects to “jabberwocky” sentences composed of opaque pseudowords, and found an early, metrically induced negativity, which peaked earlier than the classical N400 and was thus similar to that reported by Schmidt-Kassow and Kotz (2009a). The authors were thus able to lend additional support to an early meter-related negativity, which is distinct from the N400-like negativities reported above, and not directly tied to lexical access or semantic integration.

Speech rhythm has also been shown to play a role in attentional processes. Wang, Friedman, Ritter, and Bersick (2005) examined brain responses to deviant syllables in disyllabic speech sounds which subjects were instructed to ignore, and found that a change from voiced consonants to the corresponding unvoiced consonants always elicited a mismatch negativity response, but a P3a response (a member of the P300 family of components related to attentional engagement/orientation; Comerchero & Polich, 1999) only when the deviant syllable was stressed rather than unstressed, regardless of its temporal position in the item. As the subjects were instructed not to attend to the speech sounds, Wang et al. suggest that prosodic information, unlike temporal information, serves to capture attention in speech analysis. In line with such a view, Schmidt-Kassow and Kotz (2009b) found a P600 in response to slight metric deviations when subjects were instructed to focus on meter, but not when subjects were instructed to focus on grammatical structure. This finding resonates with a P600 response to metrical incongruities observed by Magne et al. (2007), which was present only when the task was explicit towards prosody rather than semantics.

The above findings lead to predictions about the ERP responses likely to be elicited by incongruous compound and phrasal stress. Below, we briefly introduce the design of the present study before discussing explicit predictions derived from these previous studies of speech rhythm/meter.

1.2 The present study

The present study employed minimal pairs of the type used by Vogel and Raimy (2002) and Vogel et al. (2009). Minimal stress pairs afford a unique opportunity for investigating contrastive stress patterns: brain responses to *identical* sets of auditory stimuli, under congruent and incongruent contexts, can be compared. To this end, we recorded continuous EEG while subjects participated in a violation paradigm in which utterances were either congruent or incongruent with a previously presented visual stimulus, as a function of the stress pattern used to label the depicted object. The visual stimulus consisted of a single image (depicting an item corresponding to only one stress pattern) in each trial. We used 44 pairs of segmentally identical (but prosodically distinct) phrases and compounds (e.g., hot dóg vs. hótdog) as test items. In experimental trials, the image (e.g., a green-colored house) established context and was followed by an utterance featuring the test item with either the congruent (green hóuse) or incongruent (gréenhouse) stress pattern, for a 2 (stress) × 2 (congruency) within-subjects design. Participants indicated (with the press of a button) whether the item depicted was named correctly. With EEG time-locked to the onset of the test item, ERPs were calculated separately for each stress pattern as the difference between congruent and incongruent trials.

Thus, a significant effect of congruency would show that prosodic mismatch with the visual context was detected in incongruent trials, suggesting sensitivity to the compound/phrasal stress distinction. If a post-perceptual bias drives the preference for compounds observed by Farnetani et al. (1988), Vogel and Raimy (2002), and Vogel et al. (2009), we would expect to find a significant brain response to the incongruent use of phrasal stress to describe images depicting compounds, and of compound stress to describe images depicting phrases, even if subjects' behavioral responses do not indicate explicit awareness of the incongruity. If, on the other hand, the compound preference stems from poor perceptual sensitivity to the distinction, we would not expect to observe a significant response to prosodic incongruity.

Regarding the role of the compound/phrasal stress distinction in online sentence processing, a number of explicit predictions can be derived from the literature. As Magne et al. (2007) found that misplaced stress accents elicited an N400, suggesting disrupted access to word meaning, we might expect a similar brain response to the incongruent use of compound or phrasal stress; if the stress contrast assists in semantic processing, a metrical incongruity may lead to a disruption which would be reflected at the scalp level by a component such as the N400. Given that a number of studies have demonstrated P600 responses to metrical incongruities (e.g., Magne et al., 2007; Marie et al., 2011; Schmidt-Kassow & Kotz, 2009a, 2009b), suggesting that metric cues interact with other information in a later integrational stage, we might expect the incongruent use of either stress pattern to drive a similar effect, given the bearing of this particular stress contrast on both semantics and phrase structure. A further possibility is that misplaced stress in incongruent trials will engage subject attention, leading to an orientation response which would be reflected by a P300-like component (as in Wang et al., 2005).

2 Method

2.1 Participants

Twenty-five University of Delaware undergraduates were recruited and received course extra-credit in exchange for their participation. All participants signed informed consent forms and completed questionnaires on language, education, and health background. Five subjects were excluded due to incomplete recording resulting from equipment malfunction (2) and experimenter error (3). Of the remaining 20 participants, 17 were female, as the majority of students enrolled in the course from which our participants were drawn were female. The mean age was 19 years (range 18–20 years). Three women were left handed; results for these subjects were included in light of studies reporting left-hemisphere language dominance in a high percentage of left-handers (e.g., Knecht et al., 2000). All subjects were native speakers of American English and reported normal hearing and normal or corrected-to-normal vision.

2.2 Stimuli and design

The experiment consisted of 176 experimental trials spread equally across four conditions (congruent compound, congruent phrasal, incongruent compound, incongruent phrasal), in addition to 112 filler trials. The image presented in a given trial (e.g., a green-colored house) established context and was followed by an utterance featuring the test item with either the congruent (green hóuse) or incongruent (gréenhouse) stress pattern, for a 2 (*stress*) × 2 (*congruency*) within-subjects design. Thus, an *incongruent compound* trial would feature an image related to a phrase (such as the green-colored house used as an example above), followed by an utterance in which the corresponding compound (in this instance, “gréenhouse”) was named. The opposite was the case for *incongruent phrasal* trials, which featured images related to compounds (e.g., a glass building with plants growing inside), followed by an utterance in which the corresponding phrase (green hóuse) was named. Test items were 44 pairs of phonetically identical but prosodically distinct phrases and compounds. In addition to the 16 minimal stress pairs used by Vogel et al. (2009), we used 28 additional pairs constructed for the present study.[1] Most of the test items were bisyllabic (33), or trisyllabic with two syllables in the first word (8), though two of the items had four syllables (two in each word). All compounds were uncontroversially stressed on the first member in American English.

Filler items consisting of semantically related pairs of phrases and compounds were included, as a means both of masking the purpose of the study and of gauging the level of subject attentiveness to the task. Both congruent and "incongruent" filler trials were included. In congruent filler trials, the item depicted in the image was named correctly; in incongruent filler trials, a random item was named which was semantically unrelated to the image (see Appendix C for a complete list of filler items). Because the incongruent trials involved items which were blatantly unrelated to the visual context, they provided an appropriate means to gauge subject attentiveness (i.e., an attentive subject would be expected to score at or near 100% on incongruent filler items, despite their performance on incongruent experimental trials). The semantically unrelated items used in incongruent filler trials did not differ from the images they were paired with in any systematic way.

Every utterance used in the study was of the form: "*This is the* [test item]." In the case of plural test items (3), the frame "*These are the* [test item]" was used. Minimal stress pairs used in the experimental condition are listed (as phrases only) in Appendix A.

2.2.1 Trial structure. Each trial began with the presentation of the visual stimulus which was followed by the utterance after 3 s. The visual stimulus was displayed continuously, overlapping with the auditory stimulus, and persisted until a behavioral response was registered (through a Serial Response Box, described below). Subjects responded by pressing one of two buttons, indicating whether or not the subject felt the item depicted in the image had been named appropriately on that trial. Immediately following the response, a feedback screen appeared indicating the subject's response choice (thereby reminding subjects which button they had pressed). Importantly, no feedback regarding the correctness of the response was given, and at no point did subjects encounter an orthographic representation of the speech stimulus. If the subject did not respond within 5 s of the auditory stimulus offset, no response was logged and a screen briefly appeared requesting faster response on subsequent trials. There was an inter-trial interval of approximately 3 s, beginning with the subject response on the previous trial. An example trial for each of the four experimental conditions is given in Appendix B.

2.2.2 Speech stimuli. In order to ensure a high degree of uniformity and avoid potential confounds due to inconsistencies in natural speech, synthetic speech stimuli were used. Auditory stimuli were developed using the ModelTalker TTS system (Yarrington et al., 2008), a concatenative synthesizer which allows control of timing and intonation. Thus, fundamental frequency and timing effects associated with pitch and phrase accents were highly consistent across the stimuli. Oscillogram and pitch contours for two utterances featuring the same test item are shown in Figure 1, illustrating the prosodic difference between compound and phrasal stress in the auditory stimulus set.

For a detailed analysis of the prosodic differences between compounds and phrases in speech synthesized with the ModelTalker TTS system, see Vogel et al. (2009).

2.3 Procedure

Following setup for EEG recording, participants were seated in a comfortable armchair in a sound- and electrically-shielded booth, facing a computer screen and speakers at an approximate distance of 1 m. Stimulus presentation and behavioral response collection were controlled by PC using E-Prime software and a Serial Response Box from Psychology Software Tools (Schneider, Eschman, & Zuccolotto, 2002).

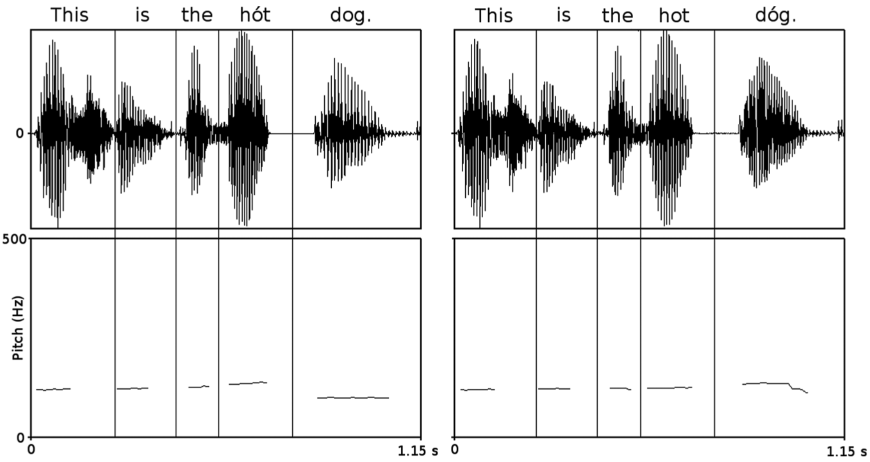

Figure 1. The left panels show the oscillogram and pitch contours for an utterance used in both the congruent and incongruent compound conditions. The right panels show the oscillogram and pitch contours for the corresponding utterance used in the phrasal conditions.

After a brief practice/task-familiarization session (lasting approx. 3 min), subjects were instructed not to blink or move unnecessarily during auditory stimulus presentation. Importantly, subjects were not informed of the purpose of the experiment or asked to attend to stress or prosody in the stimuli. Rather, they were instructed (through initial on-screen instructions) to attend to whether the correct item was named and if it was “pronounced” correctly. Subjects were instructed to press a specific key on the Serial Response Box if they felt that the test item had been named correctly and a different key if they felt it had been named incorrectly. Testing consisted of a single session divided into four blocks of 72 trials, each lasting approximately 12 minutes and followed by a short break. Impedance was rechecked (and, when necessary, electrodes were rehydrated) between the second and third trial blocks.

All subjects received a single exposure to each possible stress/congruency combination for each minimal pair (i.e., 4 for each pair). In order to control for the effects of seeing an image twice during the test session, and to minimize recognition of the purpose behind the study, the order of stimulus presentation was quasi-randomized with the constraint that half of the subjects received a specific half (50%) of the images paired with a congruent stress pattern during the first half of the session. The other half of the subjects received an incongruent stress pairing for the same images the first group received with the congruent stress pattern during the first half of the session, and vice-versa. Thus, the specific half of the images that were initially encountered in congruent trials was counterbalanced between two groups.[2] An additional constraint was that each test item was used only once in each of the four trial blocks (thus, each of the four stress/congruency combinations for each test item appeared in a different trial block).

The entire experimental session, including setup for electrophysiological recording, took approximately 1.5 hrs.

2.4 Electrophysiological recording

Continuous EEG activity was recorded with an Electrical Geodesics 300 system, using a 128-channel Geodesic Sensor Net (Tucker, 1993) of Ag/AgCl plated electrodes housed in electrolyte-soaked sponges, referenced to Cz. Data were recorded with a bandpass of .1–100 Hz and digitized at 250 Hz. Electrode impedances were kept below 50 kΩ (cf. Ferree, Luu, Russell, & Tucker, 2001). After recording, the continuous EEG was segmented into 1200 ms epochs, time-locked to the onset of the critical word(s),[3] using a 200 ms pre-stimulus baseline and a 1000 ms segment time. Following artifact decontamination (described in the next sub-section), data were baseline corrected on the 200 ms pre-stimulus period and referenced to the average voltage, which is well suited to high-density EEG (Dien, 1998; Nunez & Srinivasan, 2006).[4] ERPs were calculated separately for each stress pattern as the difference between congruent and incongruent trials.

2.4.1 Artifact decontamination. The 176 experimental trials per subject were submitted to an automated artifact decontamination procedure using *Netstation* software. A single channel in an epoch was marked as bad if fast average amplitude exceeded 200 μV, if differential amplitude exceeded 100 μV, or if it had zero variance. Channels marked as bad in over 20% of trials were considered bad in *all* trials. Trials containing more than 10 bad channels were excluded. When surrounded by channels with good data, bad channels were deleted and replaced using spherical spline interpolation. The data were then submitted to a second automated procedure which performed independent component analysis (Bell & Sejnowski, 1995) and automatically subtracted eyeblink components that correlated at r = 0.9 or greater with an eyeblink template (Dien, 2010). After both procedures, less than 11% of trials had been excluded, evenly distributed across conditions.

2.5 Statistical analysis

2.5.1 Analysis of electrophysiological data. Repeated-measures ANOVAs were performed on unfiltered mean amplitudes relative to baseline within four time windows, chosen based upon previous findings and the timing of the observable ERP responses: 0–200 ms, 200–400 ms, 400–600 ms, and 600–1000 ms. Statistical analyses covered 114 electrodes distributed among 9 regions of interest. Following the recommendations of Dien and Santuzzi (2005), we grouped lateral electrodes using three spatial factors: *anteriority* (anterior vs. posterior), *laterality* (left vs. right hemisphere, excluding midline electrodes), and *dorsality* (superior vs. inferior). Figure 2 shows the resulting 8 electrode sets. As the midline electrodes were necessarily excluded in tests involving the lateral electrodes, separate tests were also performed for midline electrodes, using *anteriority* (anterior vs. posterior midline) as a spatial factor.

Repeated-measures ANOVAs were conducted for the midline and lateral electrodes separately for each of the four time intervals. ANOVAs for the lateral electrodes included all 4 experimental conditions in a six-way factorial design based upon the factors *stress* (2: compound vs. phrasal), *congruency* (2: congruent vs. incongruent), *anteriority* (2: anterior vs. posterior), *laterality* (2: left vs. right hemisphere), *dorsality* (2: superior vs. inferior), and *block* (2: first vs. second half of the experiment), with subject as a random factor. The factor *block* was included to test for potential repetition effects, as images presented during the second half of the experiment had been presented during the first half, though paired with a different stress pattern (the counterbalanced design is described above). ANOVAs for the midline included the 4 experimental conditions in a four-way factorial design based upon the factors *stress*, *congruency*, *anteriority*, and *block*. Mean amplitudes per time window were calculated as the average amplitude over all electrodes in a given region of interest and used as the dependent measures in the ANOVAs. Analyses of specific regions were only conducted when significant interactions between condition and spatial factors were found for a particular time window.

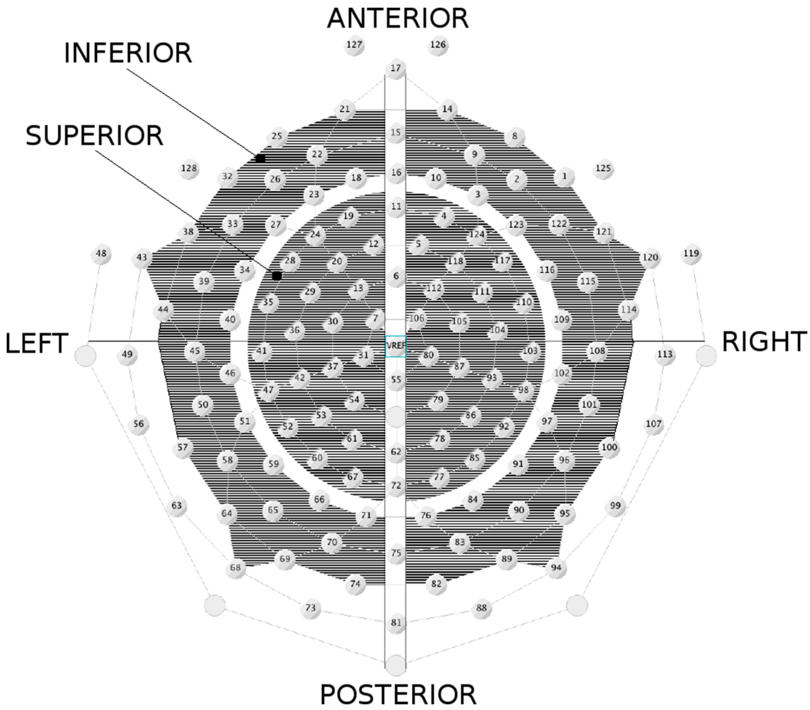

Figure 2. Electrode sets resulting from three spatial factors: anteriority (anterior vs. posterior), laterality (left vs. right), and dorsality (superior vs. inferior).

2.5.2 Analysis of behavioral data – signal detection theory analysis. In order to gain measures of subject sensitivity and response bias purely on the basis of the behavioral data, we carried out a signal detection theory (SDT) analysis (as described in Macmillan and Creelman, 1991). For congruent trials, a correct response (i.e., one indicating that the utterance matched the image) was coded as a "hit," whereas an incorrect response (i.e., one indicating that the utterance and image did not match) was coded as a "miss." For incongruent trials, a correct response (indicating a mismatch between the utterance and image) was coded as a "correct rejection," whereas an incorrect response (indicating a match between the signal and image) was coded as a "false alarm." Thus, we conceived of the "signal" as congruence between the stress pattern in the utterance and the item depicted in the image, and incongruence as "noise."

Discriminability indexes (*d'* – also known as the *sensitivity index*; a higher *d'* indicates that a signal is more easily detected) and criterion scores (a lower criterion score indicates a less conservative response pattern, i.e., greater bias towards indicating a match) were calculated for each subject on the basis of hit and false alarm rates. Following the standard method (cf. Macmillan and Creelman, 1991), *d'* scores were calculated as the difference between the z-transforms of the hit and false alarm rates (thus, a subject with identical hit and false alarm rates would receive a *d'* score of 0). Criterion scores were calculated as the negative average of the z-transforms of the hit and false alarm rates.

Importantly, two sets of *d'* and criterion scores were calculated for each subject: one set for trials in which the image indicated a compound, and one set for trials in which the image indicated a phrase. Thus, a significant difference in *d'* scores across image type would indicate greater subject sensitivity to one of the stress patterns, while a significant difference in criterion scores would indicate a response bias towards either the compound or the phrasal interpretation (i.e., subjects would be more likely to indicate that the image matched the context when a certain type of image appeared).

3 Results

3.1 Behavioral results

Behavioral results replicated previous findings of significantly greater accuracy for compound stress. Mean accuracy was 89% (*SD* 7.2%) for congruent compounds, compared to 72% (*SD* 25.6%) for congruent phrases. In incongruent trials, there was a tendency to indicate (incorrectly) that the test item matched the image; subjects responded correctly to only 32% (*SD* 27.7%) of incongruent compounds and 13% (*SD* 11.8%) of incongruent phrases. A two-way repeated-measures ANOVA, performed on logit-transformed proportions, confirmed significant main effects of stress, $F(1, 19) = 12.56$, $p < 0.01$, and congruency, $F(1, 19) = 45.17$, $p < 0.001$, with no significant interaction between stress and congruency, $F(1, 19) = 2.41$, $p = 0.14$.

Mean accuracy for the filler trials was 99% (*SD* 1.7%), indicating a high level of subject attentiveness throughout the experiment.

3.1.1 Signal detection theory analysis. The signal detection theory (SDT) analysis of the behavioral data was carried out to determine whether the pattern of greater accuracy with utterances featuring compounds stemmed from differences in sensitivity (i.e., differences at the sensory level), or from differences in response bias (i.e., differences at a higher level of decision making). Thus, the SDT analysis provided a means to use the behavioral responses as an additional complement to the electrophysiological data.

For images indicating a compound, the average subject d' score was 0.076 and the average criterion score was -1.315; for images indicating a phrase, the average d' score was 0.291 while the average criterion score was -0.771. Further analysis revealed a response bias towards the compound interpretation of test items; criterion was significantly lower when compound-congruent images set the context, $t(19) = -3.23$, $p < 0.01$. Criterion is an inverse measure of subject willingness to indicate that a signal was present in an ambiguous situation. Thus, the significantly lower criterion in this instance indicates a greater bias towards indicating that the image matched with the auditory stimulus when a compound-related image was present. However, no significant difference in sensitivity to each stress pattern was indicated, as the discriminability indexes (d' scores) did not differ significantly across the two stress patterns, $t(19) = -1.61$, $p > 0.10$.

3.2 ERP results

As subjects responded correctly to only 13% of incongruent phrasal trials, and 32% of incongruent compound trials, a separate analysis of ERPs based on correctness of response was not feasible. Furthermore, a primary motivation for the experiment was to examine whether there was perceptual sensitivity to the stress distinction using a method which did not rely on subjects' behavioral responses. Therefore, we included all trials not removed by the automated artifact detection procedures in our analysis.[5]

Inspection of the grand average waveforms revealed that voltages for both of the incongruent conditions were more negative than for congruent conditions from 400–1000 ms across the left hemisphere electrodes. This effect was most prominent at anterior electrodes and began slightly earlier (around 300 ms) for incongruent phrases, with a somewhat broader scalp distribution relative to compounds. Incongruent compounds appeared to elicit an earlier posterior negativity, peaking around 400 ms (similar to the N400 in timing and scalp topography), while a late posterior positivity (600–1000 ms) was observed for the incongruent phrases. This posterior positivity was characteristic of the P600 in both timing and scalp topography. Figures 3 and 4 provide representative channels showing ERPs for the compound and phrasal conditions, respectively. Below, we report all significant main effects of (and interactions involving) *congruency*.

Analysis of the 200–400 ms time window yielded a significant four-way interaction between the factors *stress*, *congruency*, *laterality*, and *dorsality*, $F(1, 19) = 8.46$, $p < 0.01$, stemming from a negative response to incongruent compounds, relative to the other conditions, which was most prominent at right superior electrode sites, with a corresponding positive inversion which was most prominent across left inferior electrode sites. As the interaction involved the factor *laterality*, this was followed by separate analyses for each hemisphere, which yielded a significant *stress* × *congruency* × *anteriority* interaction, $F(1, 19) = 4.60$, $p < 0.05$, for the right hemisphere. Analyses for both right hemisphere quadrants revealed a *stress* × *congruency* interaction, $F(1, 19) = 7.82$, $p < 0.05$, across the posterior quadrant, stemming from the greater negativity in response to incongruent compounds.

ANOVAs for the 400–600 ms time window revealed a significant two-way interaction between *congruency* and *laterality*, $F(1, 19) = 9.77$, $p < 0.01$, stemming from a broad left-lateralized negative response to both stress patterns in incongruent trials, relative to congruent trials, along with a corresponding right-lateralized inversion. There was also a significant *stress* × *congruency* × *block*

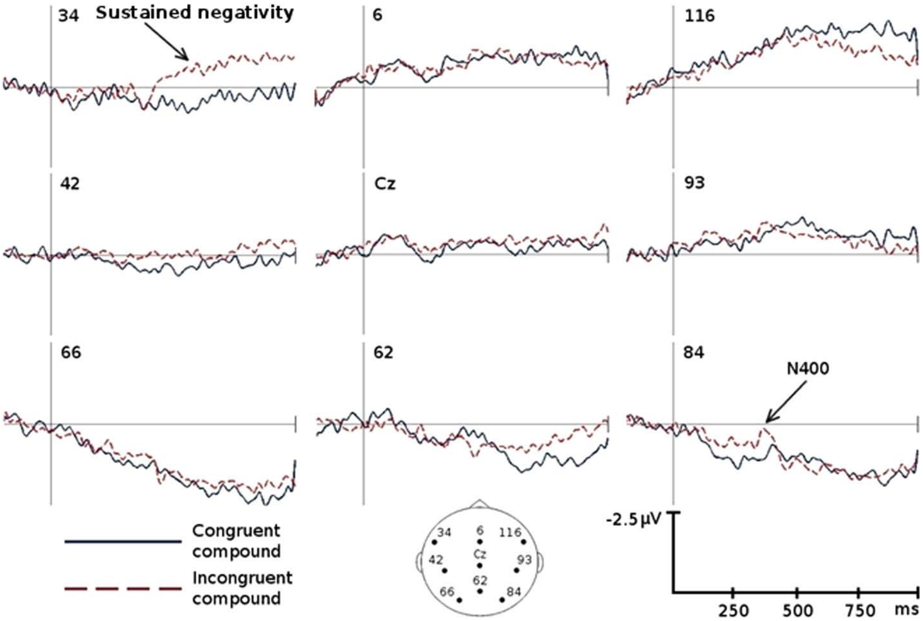

Figure 3. ERPs elicited by congruent (solid line) and incongruent (dashed line) compound stress, low-pass filtered at 30 Hz (for display purposes only). Nine representative channels (symmetrical across the hemispheres) are displayed.

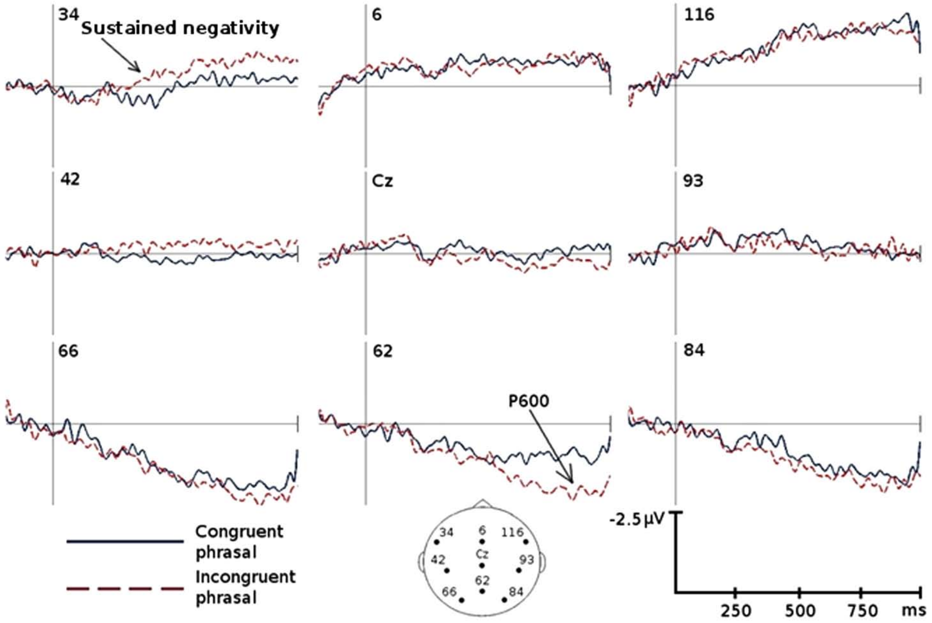

Figure 4. ERPs elicited by congruent (solid line) and incongruent (dashed line) phrasal stress, low-pass filtered at 30 Hz (for display purposes only). Nine representative channels (symmetrical across the hemispheres) are displayed.

interaction, $F(1, 19) = 4.56$, $p < 0.05$, due to more negative voltages for incongruent compounds, relative to congruent compounds, during the first half of the experiment. Separate analyses for each hemisphere revealed a significant main effect of *congruency* for both the left-lateralized effect, $F(1, 19) = 8.99$, $p < 0.01$, and its right-lateralized inversion, $F(1, 19) = 7.78$, $p < 0.05$, along with a four-way *stress* × *congruency* × *dorsality* × *block* interaction, $F(1, 19) = 5.62$, $p < 0.05$, across the left hemisphere, due to a more negative response to incongruent (relative to congruent) compounds at superior electrode sites during the first half of the experiment.

Analysis of the lateral electrodes in the 600–1000 ms time window revealed a significant interaction between *congruency* and *laterality* once more, $F(1, 19) = 4.43$, $p < 0.05$, again driven by the same effect, but neither the effect nor its inversion reached significance in separate analyses performed for each hemisphere. Analysis of the midline electrodes yielded a *stress* × *congruency* × *anteriority* interaction, $F(1, 19) = 6.77$, $p < 0.05$, due to more positive posterior voltages for incongruent phrases, relative to the other conditions. Separate analyses for anterior and posterior midline electrodes revealed a significant *stress* × *congruency* interaction, $F(1, 19) = 6.99$, $p < 0.05$, along the posterior midline, again due to more positive voltages for incongruent phrases, relative to other conditions.

3.2.1 Summary of main electrophysiological results. In summary, a sustained negativity was observed across the left hemisphere for both stress patterns when incongruent with the image. This effect was most prominent at anterior electrode sites, though there was no statistically significant interaction involving anteriority. For phrases, the effect had a broader scalp distribution in addition to beginning slightly earlier (see Figure 5). Despite these differences, the

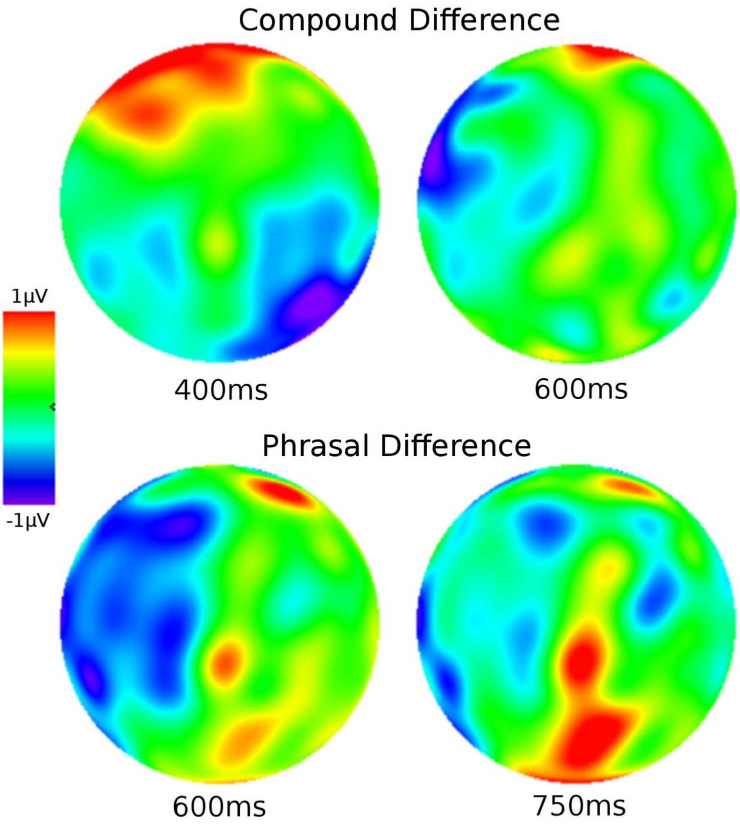

Figure 5. Images depicting the scalp topography (overhead view) of the difference waves between incongruent and congruent conditions for compounds (top) and phrases (bottom). Each of the four components described in the ERP results summary is visible: the N400 for incongruent compounds (top left), the left-lateralized negativity for both incongruent conditions (top right [compound] and bottom left [phrasal]) and the P600 for incongruent phrases (bottom right).

effect was significant during the 400–600 ms time interval for both stress patterns, as indicated by a main effect of *congruency* revealed by the hemispheric analyses. A right-lateralized centro-parietal negativity, characteristic of the N400 in both timing and scalp topography (see discussion section), was observed for compounds, but not phrases, when incongruent with the image. The statistical significance of this effect is reflected by the *stress* × *congruency* interaction in the 200–400 ms time interval in the right posterior quadrant. A late posterior positivity, characteristic of the P600 in both timing and topography, was observed for incongruent phrasal stress (but not incongruent compound stress). Analysis of this effect yielded a significant *stress* × *congruency* interaction along the posterior midline during the final time interval (600–1000 ms). Figure 5 depicts the scalp topography of each effect described in the above summary, at representative time points.

4 Discussion

The present study was conducted to determine (1) whether the apparent preference for compounds observed in previous studies stems from poor perceptual sensitivity to the compound/phrasal stress distinction, or whether it arises from a post-perceptual bias, and (2) whether electrophysiological evidence could be gained in support of a specific role for the compound/phrasal stress contrast in sentence processing. The present results help to address both questions.

Behavioral results replicated previous findings of a preference for the compound interpretation of ambiguous strings (Farnetani et al., 1988; Vogel & Raimy, 2002; Vogel et al., 2009). Interestingly, subjects had an overwhelming tendency to indicate (incorrectly) that the utterance matched the image in incongruent trials. While the possibility remains that this pattern stems from the use of synthetic speech stimuli (i.e., subjects explicitly or implicitly attributed any prosodic incongruity to imperfections in the speech synthesis), this remains unlikely, as greater accuracy was observed for compound trials (both congruent and incongruent) than for phrasal trials, a finding consistent with the behavioral results of previous studies utilizing natural speech stimuli (Farnetani et al., 1988; Vogel & Raimy, 2002; Vogel et al., 2009).

The signal detection theory (SDT) analysis of the behavioral results provides an avenue through which to explore sensitivity to each stress pattern, as well as the possibility of bias towards a given interpretation of ambiguous strings, solely on the basis of the behavioral data. The finding of a significantly lower criterion score when the visual context was set by images depicting compounds indicates a response bias towards the compound interpretation of test items. The lack of a significant difference between discriminability indexes for compound- and phrase-related images is consistent with the claim that subjects were equally sensitive to both stress patterns.

The very same compounds and phrases elicited different brain responses as a function of their congruence with the visual context. The SDT analysis of the behavioral results fits particularly well with the electrophysiological data. A left-lateralized sustained negativity was observed for both incongruent compounds and incongruent phrases, in addition to an N400-like negativity for incongruent compound stress, and a P600-like positivity for incongruent phrasal stress. As subjects responded correctly to only 32% of incongruent compounds and 13% of incongruent phrases, indicating (incorrectly) that the test item matched the context, the significant brain response to the incorrect use of each stress pattern suggests that a post-perceptual bias drives the preference for compounds observed in previous work. Had this pattern stemmed from poor perceptual sensitivity to the compound/phrasal stress distinction, we would not have expected to observe a significant electrophysiological response for trials in which the subject failed to indicate that a stress rule had been violated (which was the case for 87% of the incongruent phrasal trials and 68% of the incongruent compound trials).

Strong electrophysiological evidence for discrimination of stress is striking, given poor performance on the behavioral task. Friedrich et al. (2001) found a similar conflict between electrophysiological evidence for discrimination between pitch contours which were either congruous or incongruous with the expected stress pattern of specific words, and poor behavioral performance on a simultaneous stress evaluation task. Friedrich et al. conclude that stress information is processed automatically, whereas an explicit evaluation of stress requires higher-level controlled processes of a sort not usually involved in online spoken word recognition. This interpretation is consistent with the pattern of results observed in the present study.

4.1 Implications for online comprehension

The electrophysiological data are also well suited to exploring the role played by the compound/phrasal stress distinction in sentence comprehension, given the nature of the ERP components observed. In what follows, we discuss the further implications of the observed ERPs for online comprehension.

4.1.1 Left-lateralized sustained negativity. The left-lateralized negativity observed in response to both incongruent compound and phrasal stress was significant during the last two time windows (from 400–1000 ms), as reflected by significant interactions between *congruency* and *laterality*. A significant main effect of *congruency* was observed from 400–600 ms in separate hemispheric analyses. Though the negativity appeared to be most prominent at anterior electrode sites, the lack of significant interactions involving the factor *anteriority* suggests that the negativity is a whole-hemisphere effect. The slow drift of this negativity, combined with its broad scalp distribution, gave it an appearance similar to the contingent negative variation (CNV; Walter, Cooper, Aldridge, McCallum, & Winter, 1964) which has been observed in previous studies of metrical processing (Domahs, Wiese, Bornkessel-Schlesewsky, & Schlesewsky, 2008; Magne, Astésano, Lacheret-Dujour, Morel, Alter, & Besson, 2005), and with a left-lateralized scalp distribution in previous studies of phonological processing (Spironelli & Angrilli, 2006; Rugg, 1984).

The CNV has traditionally been identified with anticipation processes, and is often found in the interval between two associated stimuli when the second cues a response (Teece, 1972; Walter et al., 1964). It has also been argued to reflect attentional orientation (Loveless, 1979), motor preparation (Rohrbaugh & Gaillard, 1983), and increased working memory load (Ruchkin, Canoune, Johnson, & Ritter, 1995). The present CNV-like negativity is unlikely to reflect a difference in motor response preparation for incongruent trials, given poor subject performance on the behavioral task in incongruent trials (i.e., a high propensity to indicate that the auditory stimulus matched the visual context). It remains plausible, however, that the effect may reflect differences in arousal, attention, or memory load.

Two previous studies, which found CNV-like negativities in response to metrical violations, are highly relevant to the interpretation of the present effect. Magne et al. (2005) observed a CNV-like component in response to pragmatically incongruous focal accents in sentence-final position, which the authors interpret as reflecting anticipation processes stemming from prior prosodic violations (a sentence-final focus violation always meant that a preceding focus violation was heard sentence medially). Along similar lines, Domahs et al. (2008) observed a sustained negative deflection in response to misplaced stress in trisyllabic words correctly stressed on the antepenultimate syllable. Subjects were visually presented with the critical word before exposure to the auditory stimulus, allowing for prosodic expectations. Thus, the authors interpreted this effect as a CNV elicited by the substitution of an initial weak syllable for a strong one, which led subjects to maintain prosodic information in working memory until a strong syllable was heard.

Consistent with the interpretation offered by Domahs et al. (2008), the CNV-like negativity elicited by incongruent compound and phrasal stress may reflect that subjects internally activated upcoming items, based on the visual context, and were more likely to maintain this information when the incoming speech signal was inconsistent. The negativity may also reflect greater ongoing comparison of the incoming speech signal with expected prosodic/phonological patterns, in which predictions for a specific stress pattern were violated; that is, items with incongruous stress information were processed more deeply. The interpretation offered by Magne et al. (2005) of their own results may also be relevant: in the case of both incongruent conditions, a prosodic violation over the first syllable was a perfect predictor of incongruence over the subsequent syllable(s), which may have led to anticipation processes contributing to a CNV-like effect at the scalp level.

However, it should be noted that the CNV-like effects observed by Magne et al. (2005) and Domahs et al. (2008) were not strongly left-lateralized, as in the current case. Previous studies of phonological processing have, however, revealed left-lateralized CNV effects during matching tasks in which phonological information is activated by visually presented words. Rugg

(1984), for instance, found that such an effect developed during the inter-stimulus interval (ISI) in a phonological matching task, which he interpreted as involving short-term memory processes which were left-lateralized due to the nature of the task. Using a similar phonological matching task, Spironelli and Angrilli (2006) found a left-lateralized CNV which formed during the ISI and had a scalp topography highly similar to that observed in the present study. This finding was in contrast to bilaterally distributed CNV responses observed for comparable orthographic and semantic tasks using an identical set of words. Though the negativity observed in the present study was elicited by the critical item (whereas in the aforementioned studies, the CNV developed during the ISI), it may be that subjects were more likely to maintain phonological/prosodic representations in memory when the auditory input was incongruous, consistent with both the interpretation offered by Rugg (1984) and (as discussed above) that of Domahs et al. (2008). It may also be relevant that in the case of the aforementioned studies finding left-lateralized CNV effects, the phonological information was activated on the basis of visual input (as in the present study).

Thus, the present sustained negativity may reflect greater maintenance of phonological/prosodic information in memory, as well as deeper processing in the form of comparisons between the incoming speech signal and expected prosodic/phonological patterns. Regardless of this interpretation, the effect clearly suggests that subjects were perceptually sensitive to violations of expectation for *both* stress patterns, despite the compound bias evident in the behavioral results, and that this sensitivity influenced the online processing of test items.

4.1.2 Right-lateralized centro-parietal negativity (N400). The centro-parietal negativity observed in response to incongruent compounds reached significance during the 200–400 ms time window in the right hemisphere, peaking just before 400 ms. The effect was thus characteristic, in both timing and scalp topography, of a well-documented electrophysiological component known as the N400, which decades of research have linked to semantic processing (see Kutas and Federmeier, 2000, for a review). An increase in the N400 is commonly observed in response to semantic incongruities, and this is widely held to indicate greater difficulties in integrating semantic information.

N400-like components have been observed in response to metrical incongruities in both spoken and written language. Magne et al. (2005) found an N400-like negativity in response to sentence-final words with pragmatically incongruous contrastive accents in French, which the authors suggest may reflect integration difficulties. Magne et al. (2007) observed a similar N400-like deflection in response to words with misplaced stress accents, which the authors interpret as reflecting disrupted access to word meaning brought about by the changes to words' metrical structures.

Recent studies have also uncovered N400-like responses to metrical incongruities during silent reading. Magne et al. (2010) found that metrically unexpected words (i.e., stressed on the second syllable instead of the first, as expected, or vice-versa) in visually presented lists of English words elicited an N400-like negativity, which the authors interpret as reflecting the impact of the unexpected stress pattern on semantic processing. Luo and Zhou (2010) found that abnormal rhythmic patterns of the verb-noun combination in visually presented Chinese sentences elicited an N400-like negativity, and that the addition of semantic incongruence enhanced this effect (in addition to other components), which the authors take to indicate that rhythmic patterns are used in semantic integration during silent reading.

It is reasonable to interpret the present N400-like effect as similar to those reported in the aforementioned studies. While the N400 is sometimes bilaterally distributed, the right-lateralized scalp distribution of the present effect is consistent with previously reported N400 effects (e.g., Astésano, Besson, & Alter, 2004; Kiefer, Weisbrod, Kern, Maier, & Spitzer, 1998; Kutas & Hillyard, 1982). The N400 has been shown to be sensitive to global, discourse-level information (e.g., van Berkum, Hagoort, & Brown, 1999) as well as visual context (e.g., Knoeferle, Urbach, & Kutas, 2011), suggesting that the present N400-like effect, given the nature of the task, may stem from incongruities between the images (depicting phrasal items) and the semantic representations activated by the spoken compounds.

Under this interpretation, the lack of an N400 response to incongruent phrases most likely stems from the frequency and plausibility of those items – they did not activate an incongruous semantic representation to the same extent as did the (relatively more frequent and more plausible) compounds, despite violating the expected stress pattern (see the discussion of the P600 effect below).

Though all phrases and compounds featured in the current study appeared in the same simple context (the sentence frame "this is the ____"), the N400-like effect is consistent with those observed in previous studies of rhythm, which have been used to argue for a role for such information in semantic processing.

4.1.3 Late posterior positivity (P600). The late centro-parietal positivity observed in response to incongruent phrases is characteristic of the classical P600 component, in both timing and scalp topography. Although the P600 has traditionally been associated with syntactic violations (e.g., Hagoort et al., 1993), it has also been observed in response to garden path sentences (e.g., Osterhout & Holcomb, 1992), as well as grammatical, non-garden path sentences in which syntactic integration is more difficult (Kaan et al., 2000). While the P600 is often viewed as reflecting a process of syntactic reanalysis and/or repair (e.g., Friederici, 1995; Osterhout & Holcomb, 1992), the component may also reflect a process of late integration (e.g., Kaan & Swaab, 2003).

A number of P600-like positivities have been observed in response to prosodic incongruities in syntactically well-formed sentences, as well as to combined prosodic/syntactic violations. Steinhauer, Alter, and Friederici (1999), for instance, observed a P600 in response to well-formed sentences with incongruous prosodic phrasing, while Eckstein and Friederici (2005) observed a P600 in response to well-formed sentences in which the final word was prosodically marked as penultimate. More directly relevant are studies involving rhythmic incongruities. Magne et al. (2007) found that misplaced stress accents in French elicited a late positivity, and Marie et al. (2011) found that musical expertise modulated this effect. Schmidt-Kassow and Kotz (2009a) found a P600 in response to metric and combined metric/syntactic violations; because the P600 effects observed for separate metric and syntactic violations were underadditive in the combined metric/syntactic condition, the authors argued that metric and syntactic cues interact in a later "integrational" stage indexed by the P600.

Following such a view, as well as that of previous interpretations of prosodically-induced P600 effects (e.g., Eckstein & Friederici, 2005), it is possible that the P600 observed in the present study reflects difficulties integrating syntactic and semantic information with incongruent prosodic information. Such a view is compatible with a model of sentence comprehension in which different information types interact in a late revision stage (cf. Gunter, Friederici, & Schriefers, 2000). Nevertheless, this interpretation alone cannot explain why incongruent compound stress did not elicit a P600.

One plausible explanation stems from properties of the stimuli themselves: compound-congruent images may create stronger predictions for upcoming items (including the compound stress pattern) than images depicting phrases, which are less frequent, less plausible, and therefore more difficult to predict. For instance, the image of a green-painted house can produce expectations for either the word *house* or the phrase *green hóuse*, while the image of a glass building containing plants may produce a more straightforward expectation for *gréenhouse*. Thus, integration may have been hindered by the violation of a stronger expectation for a specific stress pattern in the case of incongruent phrasal trials, which featured images depicting compounds.

While subjects in the present study were not explicitly instructed to attend to prosodic information, it remains possible that the P600 reflects greater awareness of the manipulation when compound-congruent images set the context. While late positivities in response to rhythmic/metric incongruities are sometimes observed only when the task is explicit towards prosody, rather than some other aspect (e.g., semantics) of the stimulus material (Magne et al., 2007; Schmidt-Kassow & Kotz, 2009b), other P600-like components have been found in response to such incongruities even when the task is not explicit toward rhythmic aspects of the stimuli (e.g., Marie et al., 2011; Schmidt-Kassow & Kotz, 2009a). Thus, in keeping with previous research suggesting that the P600 may be attention-dependent (e.g., Coulson, King, & Kutas, 1998), the present P600 may reflect explicit processing, while the observed slow negativity may reflect the violation of implicit expectations. However, the behavioral results of the present study are not straightforwardly consistent with such an interpretation: despite the presence of a P600 in response to incongruent phrases only, subjects attained higher accuracy in incongruent compound (rather than incongruent phrasal) trials. In other words, while it remains possible that the late positivity may reflect greater awareness of the stress manipulation in the incongruent phrasal condition, this is not reflected in the behavioral data.

4.2 Repetition effects

In the analyses of the ERP data, the factor *block* was included to test for potential repetition effects, as each image appeared a second time during the second half of the experiment (though paired with a different stress pattern). As there were no interactions between *block* and *congruency* tied to any of the main ERP responses discussed above, we can safely conclude that these effects are not artifacts of repetition. However, some effects of repetition were observed: analyses yielded a significant three-way *stress* × *congruency* × *block* interaction during the 400–600 ms time window, involving more negative voltages for incongruent compounds (relative to congruent compounds) during the first block of the experiment. Consistent with this finding, the follow-up analysis of the left hemisphere for this time window (which was warranted by the significant *congruency* × *laterality* interaction in the original ANOVA) yielded a significant *stress* × *congruency* × *dorsality* × *block* interaction, due to the attenuation of the negative effect for compounds being more pronounced at left superior electrode sites during the first half of the experiment.

That the sustained negativity attenuated across blocks is perhaps unsurprising, given the extent to which subject expectations for particular stress patterns may have driven an effect of *congruency* (which was observed in the hemispheric analyses for this time window). Despite the counterbalancing of materials and the inclusion of filler items, stress expectancy violations may have diminished somewhat with a second exposure to the same images and more experience with the violation paradigm. However, the fact that *congruency* emerged as a main effect in the hemispheric ANOVAs indicates that the CNV-like negativity attenuated only somewhat in response to the repetition.

5 Conclusions

The present results demonstrate significant brain responses to the incongruent use of both compound and phrasal stress, even for cases in which subjects failed to indicate (behaviorally) that the stress pattern was incongruent with the visual context. This suggests that previous behavioral findings of a preference for compounds may stem from a post-perceptual bias (as indicated by the SDT analysis), which likely stems from the greater frequency and plausibility of the compound items. Our findings may also serve to illuminate the role of the compound/phrasal stress distinction in sentence processing. Both stress patterns were clearly utilized in online comprehension, as reflected by the left-lateralized CNV-like negativity which was statistically indistinguishable across both incongruent conditions in the 400–600 ms time window. Additional components may reflect the greater frequency and plausibility of the compound items used in the study: images depicting compounds may have triggered stronger (possibly explicit) expectations for a specific stress pattern, as reflected by the P600 observed for incongruent phrases, while compound strings themselves may have produced stronger semantic representations, as reflected by the N400 observed for incongruent compounds.

It remains to be seen whether such effects would be elicited by more naturalistic stimuli: the repetitive nature of the sentence frame may have enabled subjects to make more precise predictions about the unfolding utterance, including its meter, than would have been possible otherwise. However, as an initial step towards exploring the perceptual salience and online processing of compound/phrasal stress variation, the current results are illuminating and make explicit predictions on which experimental work extending these results might be based.

Acknowledgements

We wish to thank Edith Kaan, Cyrille Magne, and an anonymous reviewer for helpful comments and suggestions. We are also indebted to Catherine Bradley and Timothy McKinnon for help with subject recruitment.

Notes

1 In gathering enough item pairs for this study, we faced an extremely difficult task, given the rarity of compounds in English which can also be plausibly depicted as corresponding to phrases in an image. Ideally, the material would have been extended in order to use a Latin Square design, but the sheer rarity of suitable material was a limiting factor. Thus, we employ each image twice (in a counterbalanced manner, described below) and provide statistical tests for potential repetition effects.

2 Below, we report statistical analyses demonstrating that results from blocks 1 and 2 are consistent with those of blocks 3 and 4.

3 We chose to time-lock EEG to the onset of the critical item for *both* compounds and phrases. Work by Friedrich et al. (2001) indicates that pitch contours are differentiated within the first syllable.

4 While average reference is well suited to high-density EEG, a great deal of language research uses an averaged mastoids reference. For this reason, we provide images of the data (three of the same channels shown in the results section) after re-reference to averaged mastoids, for comparison, in an online supplement: http://las.sagepub.com/content/56/1/23/suppl/DC1.

5 As the filler trials were included solely to mask the nature of the manipulation and gauge subject attentiveness to the task, we excluded them from our analysis of the electrophysiological data.

References

Astésano, C., Besson, M., & Alter, K. (2004). Brain potentials during semantic and prosodic processing in French. *Cognitive Brain Research*, *18*, 172–184.

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

Comerchero, M. D., & Polich, J. (1999). P3a and P3b from typical auditory and visual stimuli. *Clinical Neurophysiology*, *110*, 24–30.

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes*, *13*, 21–58.

Cutler, A. (1986). Forbear is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech*, *29*, 201–220.

Cutler, A., & Otake, T. (1999). Pitch accent in spoken-word recognition in Japanese. *Journal of the Acoustical Society of America*, *105*, 1877–1888.

Cutler, A., & Van Donselaar, W. (2001). Voornaam is not (really) a homophone: Lexical prosody and lexical access in Dutch. *Language and Speech*, *44*, 171–195.

Dien, J. (1998). Issues in the application of the average reference: Review, critiques, and recommendations. *Behavior Research Methods, Instruments, & Computers*, *30*, 34–43.

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*, 138–145.

Dien, J., & Santuzzi, A. M. (2005). Application of repeated measures ANOVA to high-density ERP datasets: A review and tutorial. In T. Handy (Ed.), *Event-related potentials: A methods handbook* (pp. 57–82). Cambridge, MA: MIT Press.

Domahs, U., Wiese, R., Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2008). The processing of German word stress: Evidence for the prosodic hierarchy. *Phonology*, *25*, 1–36.

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

Farnetani, E., Torsello, C. T., & Cosi, P. (1988). English compound versus non-compound noun phrases in discourse: An acoustic and perceptual study. *Language and Speech*, *31*, 157–180.

Ferree, T. C., Luu, P., Russell, G. S., & Tucker, D. M. (2001). Scalp electrode impedance, infection risk, and EEG data quality. *Clinical Neurophysiology*, *112*, 536–544.

Friederici, A. D. (1995). The time course of syntactic activation during language processing: A model based on neuropsychological and neurophysiological data. *Brain and Language*, *51*, 259–281.

Friedrich, C. K., Alter, K., & Kotz, S. A. (2001). An electrophysiological response to different pitch contours in words. *NeuroReport*, *12*, 3189–3191.

Friedrich, C. K., Kotz, S. A., Friederici, A. D., & Alter, K. (2004a). Pitch modulates lexical identification in spoken word recognition: ERP and behavioral evidence. *Journal of Cognitive Brain Research*, *20*, 300–308.

Friedrich, C. K., Kotz, S. A., Friederici, A. D., & Gunter, T. C. (2004b). ERPs reflect lexical identification in word fragment priming. *Journal of Cognitive Neuroscience*, *16*, 541–552.

Gunter, T. C., Friederici, A. D., & Schriefers, H. (2000). Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *Journal of Cognitive Neuroscience*, *12*, 556–568.

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge: Cambridge University Press.

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, *8*, 439–483.

Hillyard, S. A., & Picton, T. W. (1987). Electrophysiology of cognition. In F. Plum (Ed.), *Handbook of physiology* (pp. 519–584). New York: American Physiological Society.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, *15*, 159–201.

Kaan, E., & Swaab, T. Y. (2003). Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. *Journal of Cognitive Neuroscience*, *15*, 98–110.

Kiefer, M., Weisbrod, M., Kern, I., Maier, S., & Spitzer, M. (1998). Right hemisphere activation during indirect semantic priming: Evidence from event-related potentials. *Brain and Language*, *64*, 377–408.

Knecht, S., Dräger, B., Deppe, M., Bobe, L., Lohmann, H., Flöel, A., Ringelstein, E.-B., & Henningsen, H. (2000). Handedness and hemispheric language dominance in healthy humans. *Brain*, *123*, 2512–2518.

Knoeferle, P., Urbach, T., & Kutas, M. (2011). Comprehending how visual context influences incremental sentence processing: Insights from ERPs and picture-sentence verification. *Psychophysiology*, *48*, 495–506.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, *4*, 463–470.

Kutas, M., & Hillyard, S. A. (1982). The lateral distribution of event-related potentials during sentence processing. *Neuropsychologia*, *20*, 579–590.

Loveless, N. E. (1979). Event-related slow potentials of the brain as expressions of orienting function. In H. D. Kimmel, E. H. van Olst, & J. F. Orlebeke (Eds.), *The orienting reflex in humans*. Hillsdale, NJ: Erlbaum.

Luo, Y., & Zhou, X. (2010). ERP evidence for online processing of rhythmic pattern during Chinese sentence reading. *NeuroImage*, *49*, 2836–2849.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.

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

Magne, C., Astésano, C., Lacheret-Dujour, A., Morel, M., Alter, K., & Besson, M. (2005). Online processing of "pop-out" words in spoken French dialogues. *Journal of Cognitive Neuroscience*, *17*, 740–756.

Magne, C., Gordon, R. L., & Midha, S. (2010). Influence of metrical expectancy on reading words: An ERP study. *Speech Prosody 2010*, 100432, 1–4.

Marie, M., Magne, C., & Besson, M. (2011). Musicians and the metric structure of words. *Journal of Cognitive Neuroscience*, *23*, 294–305.

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

Osterhout, L., & Holcomb, P. L. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*, 785–806.

Plag, I., Kunter, G., Lappe, S., & Braun, M. (2008). The role of semantics, argument structure, and lexicalization in compound stress assignment in English. *Language*, *84*, 760–794.

Rohrbaugh, J. W., & Gaillard, A. W. K. (1983). Sensory and motor aspects of the contingent negative variation. In A. W. K. Gaillard & W. Ritter (Eds.), *Tutorials in event related potential research: Endogenous components*. Amsterdam: North-Holland.

Rothermich, K., Schmidt-Kassow, M., Schwartze, M., & Kotz, S. A. (2010). Event-related potential responses to metric violations: Rules versus meaning. *NeuroReport, 21*, 580–584.

Ruchkin, D. S., Canoune, H., Johnson, R., & Ritter, W. (1995). Working memory and preparation elicit different patterns of slow wave event-related brain potentials. *Psychophysiology*, *32*, 399–410.

Rugg, M. D. (1984). Event-related potentials in phonological matching tasks. *Brain and Language*, *23*, 225–240.

Schmidt-Kassow, M., & Kotz, S. A. (2008). Entrainment of syntactic processing? ERP-responses to predictable time intervals during syntactic reanalysis. *Brain Research*, *1226*, 144–155.

Schmidt-Kassow, M., & Kotz, S. A. (2009a). Event-related brain potentials suggest a late interaction of meter and syntax in the P600. *Journal of Cognitive Neuroscience*, *21*, 1693–1708.

Schmidt-Kassow, M., & Kotz, S. A. (2009b). Attention and perceptual regularity in speech. *NeuroReport*, *20*, 1643–1647.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime reference guide*. Pittsburgh, PA: Psychological Software Tools.

Shahin, A., Roberts, L. E., Pantev, C., Trainor, L. J., & Ross, B. (2005). Modulation of P2 auditory-evoked responses by the spectral complexity of musical sounds. *NeuroReport*, *16*, 1781–1785.

Soto-Faraco, S., Sebastián-Gallés, N., & Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language*, *45*, 412–432.

Spironelli, C., & Angrilli, A. (2006). Language lateralization in phonological, semantic, and orthographic tasks: A slow evoked potential study. *Behavioural Brain Research*, *175*, 296–304.

Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience*, *2*, 191–196.

Teece, J. J. (1972). Contingent negative variation (CNV) and psychological processes in man. *Psychological Bulletin*, *77*, 73–108.

Tucker, D. M. (1993). Spatial sampling of head electrical fields: The geodesic sensor net. *Electroencephalography and Clinical Neurophysiology*, *87*, 154–163.

van Berkum, J. J. A., Hagoort, P. M., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, *11*, 657–671.

Vogel, I., Hestvik, A., Bunnell, H. T., & Spinu, L. (2009). Perception of English compound vs. phrasal stress: Natural vs. synthetic speech. *Proceedings of Interspeech 2009*, 1699–1702.

Vogel, I., & Raimy, E. (2002). The acquisition of compound vs. phrasal stress: The role of prosodic constituents. *Journal of Child Language*, *29*, 225–250.

Walter, W. G., Cooper, R., Aldridge, V. J., McCallum, W. C., & Winter, A. L. (1964). Contingent negative variation: An electrical sign of sensorimotor association and expectancy in the human brain. *Nature*, *203*, 380–384.

Wang, J., Friedman, D., Ritter, W., & Bersick, M. (2005). ERP correlates of involuntary attention capture by prosodic salience in speech. *Psychophysiology*, *42*, 43–55.

Yarrington, D., Gray, J., Pennington, C., Bunnell, H. T., Cornaglia, A., Lilley, J., Nagao, K., & Polikoff, J. (2008). ModelTalker voice recorder: An interface system for recording a corpus of speech for synthesis. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Demo Session* (pp. 28–31).

Appendix A: Target stimulus items (as phrases)

big birds*, big top, black belt, black board, black top, blue jay, cold sore, copper head, dark room, fat cat, flat bed, gold fish, green house, heavy weight, high chair, high light, high school, hot cake, hot dog, hot rod, lady bug, lazy boy, light house, mad man, orange tree, paper boys, paper weight, red coat, red head, red wood, rose buds, silk worm, sky light, soft ball, star light, tight rope, top hat, toy store, upper cut, white board, white cap, white house, yellow jacket, yellow pages

* multiple instances of the Sesame Street character, "Big Bird," were depicted in the image corresponding to the compound.

Appendix B: Example trials

Congruent phrasal trial – 3 s familiarization to the image of a green-painted house, after which the auditory stimulus is presented: "This is the green hóuse."

Incongruent phrasal trial – 3 s familiarization to the image of a blackboard (chalkboard), after which the auditory stimulus is presented: "This is the black bóard."

Congruent compound trial – 3 s familiarization to the image of a lighthouse, after which the auditory stimulus is presented: "This is the líghthouse."

Incongruent compound trial – 3 s familiarization to the image of an orange-colored tree, after which the auditory stimulus is presented: "This is the órange tree."

Appendix C: Filler items

Compounds: bathtub, classroom, fishbowl, birdhouse, jumprope, bookcase, carseat, cupcake, toothbrush, snowflake, doorknob, teapot, footprint, raincoat, sailboat, schoolbus, stopsign, tablecloth, mousehole, snowman, lampshade, bedroom, flowerpot, horseshoe, hairbrush, rainbow, candycane, cellphone, teddybear, pencilcase

Phrases: full tub, empty room, big fish, purple bird, blue rope, thick book, old car, tall cake, dark tooth, deep snow, round door, shiny pot, wet foot, green coat, large boat, green bus, striped sign, square table, pink mouse, tall man, thin lamp, big bed, red flower, brown shoe, curly hair, heavy rain, long candy, old phone, sleeping bear, yellow pencil