From Intonation to Information: an ERP study of the effects of intonation on sentence processing

Kira Mourao
The University of Edinburgh

From Intonation to Information:

an ERP study of the effects of intonation on sentence processing

Kira Mourão (nee Smyllie)

Masters Thesis

MSc in Neuroinformatics

University of Edinburgh

18.08.2006

Supervisors:

Mark Steedman	- School of Informatics, University of Edinburgh
David I. Donaldson	- Department of Psychology, Stirling University

Table of Contents

Acknowledgements

With thanks to both my supervisors for many helpful discussions and support amidst some very busy schedules. I am also grateful to everyone in the PIL lab for their help, and especially Graham MacKenzie and Lucy MacGregor, for their excellent advice on countless occasions.

Finally I would like to thank my husband for his patience, his support and his laptop.

1 Abstract

Several theories predict a link between intonation and information structure through the L+H* and H* pitch accents. In particular, Steedman [1] proposes that the L+H* LL% tune marks theme and the H* LL% tune marks rheme. In the context of ERP studies of sentence processing, the theory predicts that inappropriate accenting of theme and rheme will affect semantic processing, eliciting an N400 effect. This study experimentally investigates the prediction by varying expected information structure, pitch accent and intonational phrase boundaries in simple sentences presented to subjects, while recording their EEG responses. A small but significant effect of intonational appropriateness was found, coincident with the N400 time window. The results support Steedman's theory and set a direction for future work to quantify the topographic distribution of the effect and its relation to the N400.

2 Introduction

Intonation in English sentences can dramatically affect the listener's understanding. Consider the example below, taken from Steedman [1], where the answers to two different questions have the same wording, but different intonations, due to the context of the question. In (1) the answer to the question “Anna” is spoken with a different pitch to “Manny”; in (2) this is reversed, with the pitches on “Anna” and “Manny” switched:

(1) Q: *I know EMMA will marry ARNIE. But who will marry MANNY?*

A: *ANNA will marry MANNY.*

(2) Q: *I know EMMA will marry ARNIE. But what about ANNA?*

A: *ANNA will marry MANNY.*

Swapping the answers to the questions immediately makes the dialogues confusing, when spoken, suggesting that the intonation used in these sentences affects the listener's comprehension.

2.1 Intonation

Intonation is one of a range of acoustic features which make up the prosody of an utterance. Prosody covers the stress, duration and pitch of syllables as well as pauses between different phrases of a sentence. These can be divided into metrical and intonational components [2]: the metrical component consists of those features of the sentence which give it rhythm (changes in stress and duration across different syllables), while the intonational component consists of features affecting the pitch of the sentence (changes in pitch across different syllables).

In English the ToBI analysis system [3], based on work by Pierrehumbert [4], is commonly used to analyse the prosodic properties of sentences. In terms of the ToBI system, a sentence consists of a three-tiered hierarchy of prosodic phrases and pitch accents. At the highest level a sentence is made up of intonational phrases (IPh), which must contain one or more intermediate phrases. Intermediate phrases (ip) group words together, with each phrase consisting of a series of pitch accents on important words. Only two pitch targets are considered, high (H) and low (L), so pitch accents, marked with * on the stressed syllable, can be high (H*) or low (L*), or more complex combinations such as with leading or following pitch targets (e.g. L + H*, L* + H). At the phrase level, intermediate phrases end with a phrase tone which can be high (H -) or low (L -), and similarly intonational phrases end with a high or low boundary tone (H% and L% respectively).

Pitch tracks are a visual representation of the changes in pitch over a sentence, i.e. the fundamental frequency contour F_0, and can be used to demonstrate the prosody of a sentence. The pitch tracks for the earlier example ("Anna will marry Manny") are shown below (note that the phonology of this example has been extensively studied by Liberman and Pierrehumbert [5]):

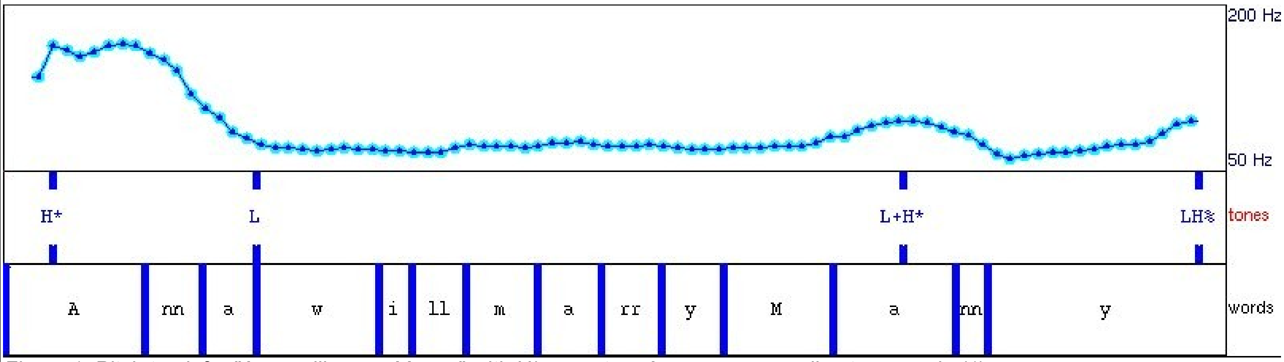

Figure 1: Pitch track for "Anna will marry Manny" with H accent on Anna, corresponding to example (1)*

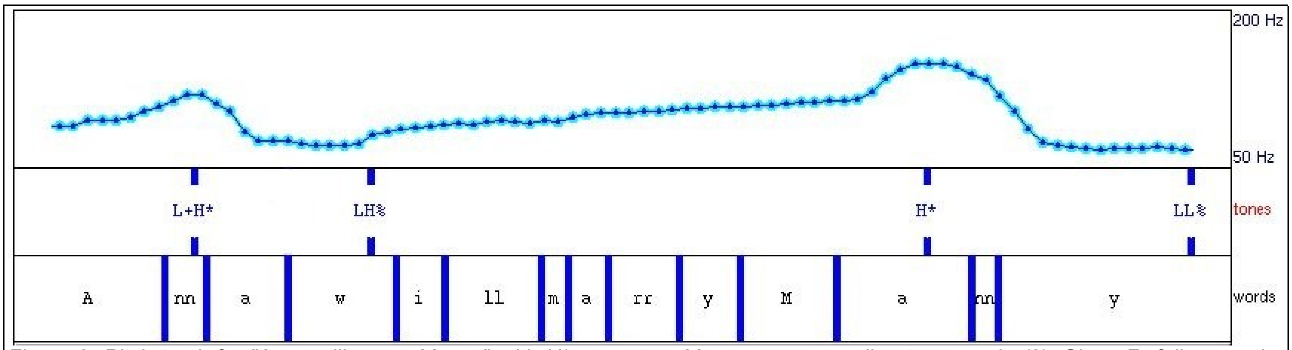

Figure 2: Pitch track for "Anna will marry Manny" with H accent on Manny, corresponding to example (2). Since F_0 falls over the course of a sentence ("declination" [6], [7]) H* is less prominent in this case than in figure 1, where the H* accent occurs at the start of the sentence.*

Here it can be seen that the H* and L+H* accents used in different positions in the same sentence can produce two distinctly different answers. According to the ToBI system, H* and L+H* are both types of "peak accent" where in both cases the peak pitch is timed to occur on the accented syllable. The timing of this peak in fact varies depending on the prosodic context – the length of the accented syllable and the value of the subsequent tone. Typically the peak will be early in long syllables especially if followed immediately by an L- phrase accent, but in short syllables with no immediate following phrase tone the peak can be late or even after the end of the syllable.

The difference between H* and L+H* accents is in the rise to the high tone. An H* peak is characterised by "at most a small rise from the middle of the speaker's voice range", whereas the L tone in L+H* indicates that the high tone rises from a low F_0 value which is not due to a preceding L* pitch accent, L- phrase accent or L% boundary tone. Additionally the rise in this case should be late in comparison to an H* accent. Because of the contextual information about the speaker's pitch range required to identify an L+H* tone, the ToBI annotation guidelines [8] stipulate that peaks at the start of a sentence (or any peak where there is some doubt) should always be marked as H*. (This makes corpora annotated in the ToBI system unreliable with respect to the L+H*/H* distinction.) Note that in the annotation above this directive has been ignored, because the tracked utterances were made in the context of a series of other utterances and it is clear from this additional context that the tones marked L+H* rise from a relatively low F_0 value.

The nature of H* and L+H* accents is disputed, both in terms of the exact definition of H* and L+H*, and also whether they are actually different. Taylor [9] demonstrates that annotators' classifications of pitch accents as H* and L+H* overlap when transformed to a different model, and questions whether they are different intonational categories. However, Steedman [1] argues that the annotation guidelines encourage miscategorisation of L+H* as H*. Furthermore, Calhoun [10] shows that the ToBI system may exclude features essential to differentiating between L+H* and H*. Her results support Ladd and Schepman's proposal [11] that L+H* and H* align with different segments, with the rise for H* occurring anywhere before stressed syllable, whereas the rise for L+H* is at the vowel of the stressed syllable. The results also suggest that pitch height on H* accents is higher and the difference between initial and final pitch of the H* accent is larger.

Comparing the pitch tracks for the previous "Anna will marry Manny" example, it can be seen that the L+H* and H* accents marked in the track also display the characteristics described by Calhoun, in terms of pitch height and initial/final difference, although the segmental alignment is less clear-cut.

2.2 Information structure and the link to intonation

In English, there is a connection between the intonation used in a sentence and its *information structure.* Information structure can be broadly defined as the organisation of the information communicated by a sentence in relation to the discourse context. It is determined by how the sentence corresponds to the foregoing discourse, by the speaker's point of view in terms of their knowledge, beliefs, intentions, expectations etc. about the sentence content, as well as by the speaker's assumptions about the hearer's point of view [12].

A number of theories have attempted to categorise the underlying components of information structure. The exact nature and definition of these components is not agreed upon, and a proliferation of terms, definitions and categorisations exists in the literature. However, a common feature across theories of information structure is a distinction, in some way, between known and unknown information. Kruijff-Korbayová and Steedman [12] point out that all theories include at least one of two major distinctions. One distinction is between the parts of an utterance which link it back to the preceding discourse, and parts which bring new information into the discourse, the former being “topic” or “theme” and the latter “comment” or “rheme” [13],[14],[15]. Another distinction is between parts of an utterance which determine the actual content of the utterance, out of the set of possibilities afforded by the context, termed “new” or “kontrast”, and those parts which do not, “given” or “background” [14]. Beyond this, theories differ both in terms of the exact definitions of these concepts, and in how they relate to each other (e.g. equivalent, orthogonal, embedded).

Information structure is realised differently in different languages; in English, this seems to be largely by intonation, although word order and syntactic structure also contribute. Unsurprisingly, given the debate over information structure, many different proposals have been made regarding how intonation relates to it. Several theories propose that a specific combination of pitch accents, phrase and boundary tones are used to differentiate between (some form of) known and unknown information. Again, the exact combination is disputed, but most suggest that the H* pitch accent is part of the tune which marks unknown information, while the L+H* accent is part of the tune which marks known information.

Jackendoff [16] was one of the earliest to propose that his 'falling' A-accent indicated focus, and his 'fall-rise' B-accent, topic. Pierrehumbert [4] related the A-accent to an H* LL% tune and the B-accent to H* LH% tune, and later Pierrehumbert and Hirschberg [17] stated that L+H* is more likely to occur in a contrastive context, that is, to indicate that the accented item should be believed, in contrast to some other possible item (similar to "kontrast" for information structure). More recent work has proposed L+H* marks known information and H* unknown information [18], [19], [20], [15].

In particular, in Steedman's theory [1], [18] sentences can be decomposed into theme and rheme components, where informally, theme is defined to be "the meaning of an utterance supposed *already* to be common ground" and rheme to be "the meaning of an utterance which it seeks to *make* common ground" [1]. Steedman claims that the theme and rheme components of a sentence are marked by distinctive tunes, and specifically identifies the ToBI L+H* LH% tune as one marking theme, and H* LL% and H* L as tunes marking rheme. This is in a wider context of a model of intonation structure based on four dimensions of literal meaning, of which the information-structural role is one. The focus of this dissertation will be on the information-structural aspect of Steedman's theory.

2.3 Experimental evidence

Experimental research on the link between intonation and information structure has been limited. Off-line studies such as the analysis of empirical data from corpora, and production/perception studies have not conclusively linked the ToBI H* and L+H* accents to information structure. For example, Hedberg and Sosa [21] tested whether H* marked focus and L+H* marked topic in a sample of spontaneous dialogue. In some support of the theory, they showed that H* and L+H* were the most commonly occurring accents, and that there were correlations between information structure and pitch accent. They were unable to show the known/unknown dichotomy suggested by the literature, although they did conclude that L+H* often marked the information structure of "contrastive focus" defined by Gundel [20].

Calhoun approached the issue by testing the accents produced on themes and rhemes, characterising them and testing the acceptability of the resulting accents in a perception study [10]. A naïve speaker was used to produce a natural-sounding set of sentences, specifically constructed so that each target word would occur as both theme and rheme in both first and second clauses of the sentences. Pitch accents on the theme and rheme, if produced, were consistently produced differently. Listeners also perceived the pitch accents differently, and preferred sentences where the theme/rheme accents were correctly placed. The position of the the theme affected the results: listeners had more difficulty detecting incorrectly placed pitch accents on sentences where the theme was positioned first (theme-rheme order is more common than rheme-theme order in English). However, theme and rheme accents could not be clearly identified as L+H* and H*, but instead were characterised by differences of relative pitch height, and the timing of the pitch rise onset.

On-line studies have mostly used EEG methods (discussed further below) although an eye-tracking study has been used to test whether H* signals new information and L+H* the presence of contrast [22] . Eye movements were tracked between two pictures whose names shared the same first accented syllable (e.g. camel, candle) where one word was new and the other was contrastive. They showed that an L+H* accent biased the listener to the contrastive word but that an H* accent did not produce a bias (although they do not discuss the possibility that the H* accent reduced a natural bias to the contrastive word). The example in this case was a contrastive theme compared with a non-contrastive rheme in different contexts, whereas the Calhoun study used a theme contrasted with a rheme in the same context, however differences between L+H* and H* were still present.

These results suggest that there is some basis for linking the H* and L+H* pitch accents to information structure, but add further weight to the dispute over the definition of these pitch accents.

2.3.1 EEG applied to intonation

In order to investigate the effects of intonation on language processing, the work reported in this dissertation uses EEG (electroencephalogram) recordings of brain activity while subjects listen to dialogues with varying intonation. EEG is a measure of the electrical activity produced by the brain, taken from electrodes positioned on the scalp of a subject. It has high temporal resolution, picking up changes in electrical activity at millisecond timescales. This compares favourably with other techniques such as fMRI, but with the trade-off that EEG recordings have poorer spatial resolution.

The event-related potential (ERP) technique is commonly used with EEG recordings. ERPs are voltage fluctuations in the brain in response to an event, such as an external sensory event, motor event or internal mental event. The changes in the electrical signal arise from the synchronous activity of populations of neurons, mainly cortical pyramidal neurons. At the level of a single pyramidal neuron, activation at a synapse causes an influx of positively charged ions, leaving the surrounding extracellular space negatively charged. As the current at the synapse flows down the dendrite towards the cell body, it dissipates across the cell membrane, resulting in positively charged extracellular space further away from the synapse. At this moment there is a dipole across the dendrite: negative charge closer to the synapse and positive charge closer to the cell body [23]. Although this tiny dipole would not be detected by a scalp electrode, pyramidal cells are uniformly oriented in the cortex, and when synchronously activated this electrical potential is summed across many thousands of neurons, allowing it to be detected at the scalp. Even though the potentials are summed over thousands of neurons, ERPs are small (voltages around 5-10μV) and individually difficult to separate from other electrical activity in the brain (voltages around 50-100μV). However since ERPs are time-locked to an external event, signal processing techniques can be used: multiple recordings can be made and averaged to produce a signal representative of the brain activity in response to the event.

In applications of ERPs to cognitive psychology, it is assumed that ERPs result from brain activities related to cognitive processes. Indeed, experimental evidence shows that manipulation of variables believed to affect various cognitive processes does correspondingly affect ERP waveforms. Where differences in ERPs are observed in response to different experimental conditions, it can then be inferred that the cognitive processes involved are also different for the different conditions. (Although note that the differences in ERPs will not necessarily correspond directly to a specific cognitive process, as the ERPs are the result of a combination of electrical signals from different areas of the brain.)

ERP components are specific, repeatable changes in ERP waveforms observed in response to experimental manipulations of cognitive processes, assumed to be generated by particular neuronal processes and related to particular cognitive functions. ERP components are usually characterised by the polarity of the waveform deflection (positive P or negative N), latency of the peak deflection with respect to the stimulus (measured in ms), and scalp distribution. A number of language-related ERP components are now well established, such as the N400 and P600 components, the left anterior negativity (LAN), early-LAN (ELAN) and the closure positive shift (CPS). P300 components, although not language-specific, may also be elicited in language-related contexts. These ERP components are described in more detail below.

The N400 is a relatively negative-going deflection over the centro-parietal area, which peaks at around 400ms after the onset of a word. This was first shown by Kutas and Hillyard [24],[25] who compared responses to visually presented sentences which ended either in a semantically congruous word (e.g. "He spread the warm bread with <u>butter</u>") or a semantically incongruous word (e.g. "He spread the warm bread with <u>socks</u>"). The elicited waveform was more negative in the incongruous case, from 200–500ms after stimulus onset, peaking at approximately 400ms. The amplitude of the N400 is inversely correlated with the degree to which the stimulus word is semantically related to the sentence context [26], suggesting that the N400 indicates semantic expectancy. Later studies have shown that the N400 effect also occurs during auditory sentence processing [27], [28], [29]. The N400 effect appears to be specific to semantic meaning, not being sensitive to physical incongruities such as typeface [24] or grammatical errors [30], but sensitive to possibly meaningful stimuli such as pictures, faces or environmental sounds [31].

The left anterior negativity (LAN) occurs in the same time frame as the N400, but typically peaks over the left anterior regions of the brain. Furthermore, this component is elicited in response to syntax violations such as morphosyntactic errors (e.g. subject-verb agreement) [32], or verb inflection errors (e.g. inflecting an irregular verb as a regular verb) [33]. The LAN often occurs after an early left anterior negativity (ELAN), considered to be an automatic response to word category errors [34]. The ELAN is situated in the same left anterior location as the LAN but peaks much earlier, at 100-250ms after the onset of the erroneous word. In contrast to the N400 component, the ELAN and LAN are clearly related to syntactic processing. These components are often followed by a much later syntax-related positive component, the P600.

The P600 is a relatively positive-going centro-parietal deflection which peaks at 600ms after stimulus onset, but in this case the typical stimulus is the onset of a word which violates the syntax of the sentence. Osterhout and Holcomb [35] compared ERP responses to sentences which were either grammatical (e.g. "The broker persuaded to sell the stock was sent to jail") or ungrammatical (e.g."The broker hoped to sell the stock <u>was</u> sent to jail.") and demonstrated that a more positive waveform was elicited in the ungrammatical case, lasting from 500-800ms after stimulus onset and peaking at approximately 600ms. As well as syntactic violations, the P600 effect also occurs for syntactically complex sentences [36], or sentences requiring syntactic reprocessing (garden path sentences) [37]. It has been suggested that the P600 component is related to higher-level syntactic processes than the LAN and early-LAN, such as syntactic reanalysis and repair, as opposed to, for example, identification of syntactic structure [34]. As with the N400, further studies have shown that the P600 can also be elicited in the auditory modality [38].

In contrast to the components discussed above, the P300 component is not specific to language. The P300 is characteristically evoked via an “oddball paradigm” where two stimuli (e.g. tones, letters, faces) are used, one much less frequently than the other (1 in 3 presentations or fewer) [39]. Relative to the ERP response for the frequent stimulus, the infrequent stimulus evokes a positive-going deflection peaking at 300ms. The P300 component is more accurately the combination of at least two separate components occurring in approximately the same time frame: the P3a component is frontally distributed and has shorter latency relative to the the P3b component, which is more parietally distributed. While the P3a component is thought to be related to attentional switching [40] , the P3b component is believed to correspond to the indexing of memory storage [41].

Finally, the closure positive shift (CPS) is a positive deflection in ERP waveforms, which corresponds to the presence of an intonational phrase boundary in comparison to no phrase boundary [42]. The CPS is a broadly distributed positivity, peaking around the parietal midline. It has very short latency, approximately 100-200ms after the intonational phrase. Even when sentences contain no linguistic information, either using nonsense words or hummed, the CPS is present, indicating that it is a purely prosodic effect [43]. However this is of particular interest as it shows that prosodic information processing occurs during semantic and syntactic processing rather than being integrated at a later stage.

Previous studies have investigated the ERP response to anomalous pitch accents, typically using simple question-answer dialogues as test cases, and varying the pitch accents in the answers. These have focused on a single pitch accent of a single kind in a sentence. Many of the studies which have been done were not carried out in English, so additionally, the results may not apply to English.

For example, Magne et al [44] used (French) dialogues constructed so that the same sentence answered two different questions (if pitch was ignored). For one question, a sentence-medial noun in the response would be the rheme, and a sentence-final noun the theme, and vice versa for the other question. (e.g. “Did he give his fiancée a ring or a bracelet?” “He gave a ring *(rheme)* to his fiancée *(theme)*.” in contrast to “Did he give a ring to his fiancée or his sister?" "He gave a ring *(theme)* to his fiancée *(rheme)*.") They then varied the pitch accents in the responses, to either accent only the theme or only the rheme, resulting in 4 test conditions - expected position (medial/final) versus actual position (medial/final) of the pitch accent. Subjects were tasked to indicate whether the sentences were prosodically correct. Magne et al identified P3a and P3b components in response to sentence-medial words, for inappropriate accents (P3a and P3b), and missing accents (P3b). They suggest the inappropriate accent induces attentional switching, hence the P3a component, while both inappropriate and missing accents are task relevant and so elicit a P3b component response. In contrast, sentence-final words with either inappropriate or missing accents evoked an N400 response in comparison to correctly accented words. Heim and Alter [45] show equivalent results for German.

Johnson et al [46] carried out a similar study in English. Test cases also varied the position and accenting of themes and rhemes, but using only one pitch accent, H*. Experimental results showed an early (100-500ms after word onset) mostly anterior negative response to a missing accent compared to the expected accent, although only on the rheme and only when in second position in the sentence. These results agree with those demonstrated by Hruska et al [47] for German, and Ito and Garnsey [48] for Japanese, although in the Japanese case the response was more positive for missing accents. Hruska's and Johnson's studies also showed an effect of information structure: the response to a theme in comparison to a rheme was a broadly distributed, bilateral positivity present at both anterior and posterior electrodes, 400-600ms after the onset of the theme or rheme.

Toepel and Alter [49] carried out two ERP experiments where subjects listened to sentences with inappropriate or appropriate intonation in German. In the first experiment they answered questions about sentences they heard, and in a second experiment made judgements about the prosodic appropriateness of the sentences. ERP signals were affected by the task, specifically negative peaks observed in response to prosodic incongruence in the first experiment only occurred for missing focus accents in the second experiment, which the authors suggest shows that "explicit structural analysis suppresses overspecified information". This has implications for comparisons between experiments, as subjects are usually tasked with either prosodic judgement tasks or delayed probe tasks, and it probably means that prosodic judgement tasks should be avoided.

2.4 The present study

Although previous studies have investigated the relationship between pitch accents and information structure, this has been limited to comparisons between the presence or absence of a pitch accent, rather than comparing two different accents. Intonational phrase boundaries in relation to information structure also have not been considered. The earlier studies suggest that online processing of prosodic information with information structure does occur to some extent, but there is no consensus on the responses to congruous versus incongruous pitch accenting. It does seem that differences between the position of theme and rheme in a sentence will affect the response [44], [46], [47], as may the task assigned to subjects [49].

The study therefore investigates the ERP responses to sentences with correct and incorrect intonation with respect to theme and rheme, in terms of Steedman's theory. As pitch accent, phrase boundary and information status are all relevant, all three factors are varied in the experimental materials. To test the theory, sentences were constructed where the intonational tunes either matched the information structure or did not. If H* LL% and L+H* LH% tunes correspond to rheme and theme then it would be expected that the tune would influence semantic processing, or be processed with other semantic cues. As discussed, semantic processing can be detected in ERP studies via the amplitude of the N400 deflection, in particular, the N400 amplitude is a measure of semantic incongruity and corresponding additional processing effort required to integrate incongruous sentence patterns. Therefore if intonational tunes relate to semantic aspects of the sentence then an N400 would be expected in the presence of anomalous accenting. Given the exploratory nature of the experiment, the data will also be analysed for other signals e.g. P300, LAN components.

3 Methods

3.1 Participants

11 participants participated in the experiment, recruited from students and staff at the University of Stirling and the wider community. Participants were right-handed native English speakers with normal hearing and vision, and with no known language, hearing or neurological disorders (by self-report). Mean age was 27.4 (range 22-33), 3 males, with participants educated to undergraduate level or above. All participants gave informed written consent for the experiment, which was approved by the University of Stirling Ethics Committee. The experiment took around 3 hours, including fitting the electrode cap, and participants were paid £15 for their time.

A further 3 participants were tested. One of these did not carry out the experiment due to difficulties with the equipment, and two were excluded from the analysis after data processing, as too few trials remained (see section 3.5).

3.2 Materials

Since the intonation examples need context in order to determine whether they make sense or not, all of the materials consisted of a dialogue. Context was established by a statement and a question, followed by a response. In the responses the following variables were manipulated:

- first occurring accent (either H* or L+H*),
- intonation (either correct or anomalous) and
- information structure (indicated by early or late phrase boundary),

giving eight test conditions.

All statements were structured as a declarative sentence preceded by “I know” while the questions took four possible forms, with each form eliciting a different information structure and first occurring accent. All responses were structured as declarative sentences of the form:

<noun3> <transitive verb> <noun4>

e.g. Anna will marry Manny

where each noun could be a proper name or was prefixed with “the”, and each verb could be past or future tense. The responses contain one theme and one rheme (both restricted to be nouns), so each response has two different pitch accents. By varying whether the given information in the question contains either noun3 or noun4, the correct response will have the H* accent on noun4 or noun3 respectively. Similarly, the phrase in the question after the *wh-* word corresponds directly to the phrasing used in the response, so by varying the question content after the *wh-* word, the information structure of the response can be varied. This creates four types of response: responses with different information structure corresponding to either early or late phrase boundaries, and responses with different intonation corresponding to either an initial H* accent or an initial L+H* accent. Figure 3 summarises the structure of the materials.

I know <noun1> <transitive verb> <noun2>	*But who/what <transitive verb> <noun4>?*	*<noun3> <transitive verb> <noun4>*	*short initial H* phrase*
	But what about <noun3>?		*short initial L+H* phrase*
	But who/what did/will <noun3> <transitive verb>?		*long initial L+H* phrase*
	But what about <noun4>?		*long initial H* phrase*
I know Emma will marry Arnie.	*But who will marry Manny?*	*Anna will marry Manny*	*(Anna) (will marry Manny)* *H* L L+H* LH%*
	But what about Anna?		*(Anna) (will marry Manny)* *L+H* LH% H* LL%*
	But who will Anna marry?		*(Anna will marry) (Manny)* *L+H* LH% H* LL%*
	But what about Manny?		*(Anna will marry) (Manny)* *H* L L+H* LH%*

Figure 3: Structure of experimental materials

In all of the intonationally correct responses, the first accented noun of the response is correctly marked either as theme (L+H*) or as rheme (H*). Intonationally anomalous dialogues were then created by swapping answers to different questions, but still maintaining the same information structure i.e. swapping early boundary H* phrase with early boundary L+H* phrase, and swapping late boundary H* phrase and late boundary L+H* phrase. In this way, acoustically identical responses make up both the intonationally correct and incorrect test cases.

This means that a test case consists of 8 declarative responses, half of which are intonationally anomalous. 37 test cases were constructed, providing 296 question/answer dialogues. In order that participants can be their own controls for different answer intonations, participants always heard every dialogue in a test case, although not consecutively, and all test cases. Example test cases can be found in the appendix – note that the nouns chosen for the test cases were specifically chosen to be easy to pitch track, using consonants l,m,n and r where possible, to avoid instability in the pitch tracks.

All test cases were recorded prior to the experiment and pitch-tracked to ensure that the intonation was as claimed. The differing intonation patterns for the different responses can be clearly seen in Figures 4 - 7. Each figure shows a plot of pitch against time for one condition of every test case, i.e. figure 4 depicts the pitch of all responses with short initial H* phrase:

A: *(ANNA) (will marry MANNY).*
H* L L+H* LH%

figure 5 depicts responses with short initial L+H* phrase:

A: *(ANNA) (will marry MANNY).*
L+H* LH% H* LL%

figure 6 responses with long initial L+H* phrase:

A: *(ANNA will marry)* *(MANNY)*

L+H* LH% H* LL%

and figure 7 responses with long initial H* phrase:

A: *(ANNA will marry)* *(MANNY)*

H* L L+H* LH%

The time length of the pitch track of each sentence was scaled to allow all the sentences to be plotted together. Since the sentences are different lengths and the contents of the sentences vary, the time courses of the pitch tracks also vary. However, the characteristic patterns of each type of sentence are clearly apparent in each plot, demonstrating the consistency of the experimental materials.

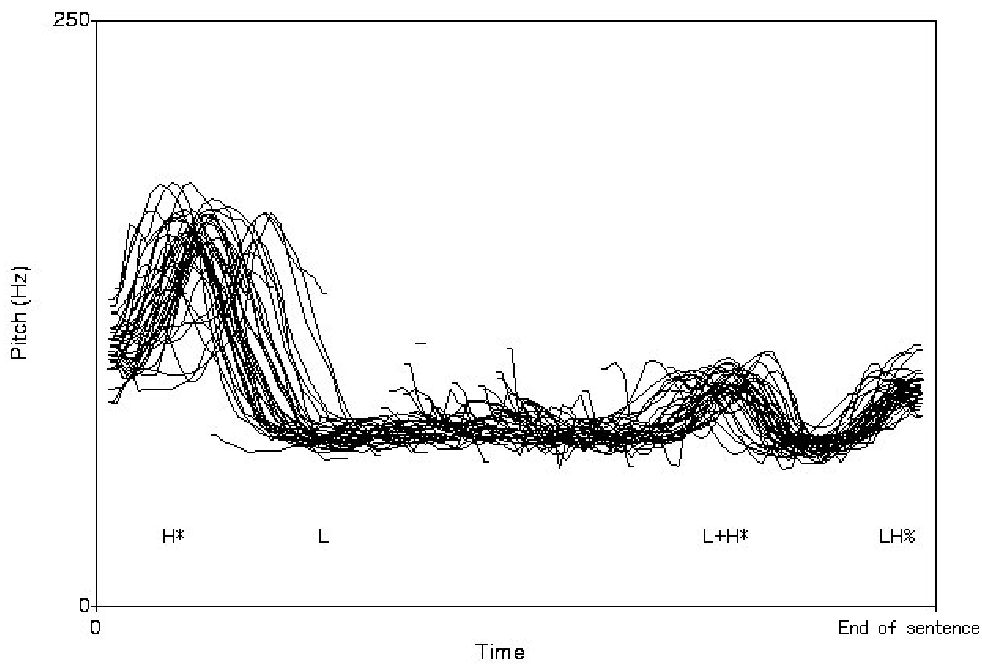

Figure 4: Pitch tracks of all responses with initial H accent and early phrase boundary. The high H* peak is followed by a drop to the L phrase boundary. The sentence ends with the lower L+H* peak followed by the L phrase accent and H% boundary tone.*

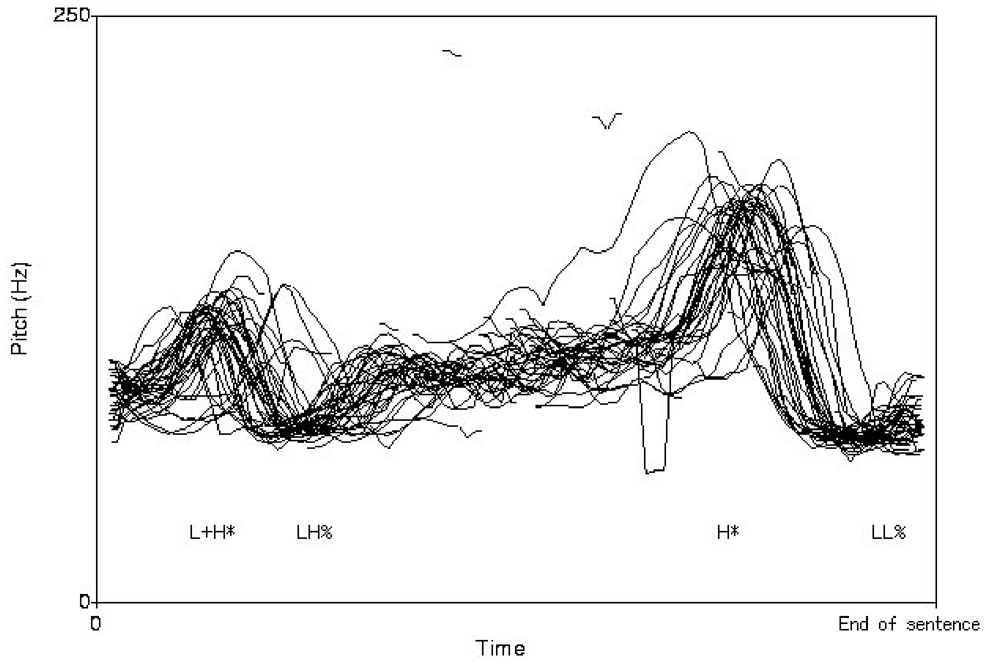

Figure 5: Pitch tracks of all responses with initial L+H* accent and early phrase boundary. The initial L+H* peak is followed by an L phrase accent and a clear rise to the H% boundary tone. The later H* peak is visibly higher than the L+H* peak and is followed by an L phrase accent and L% boundary tone, indicated by the pitch not returning to the previous 'high' level.

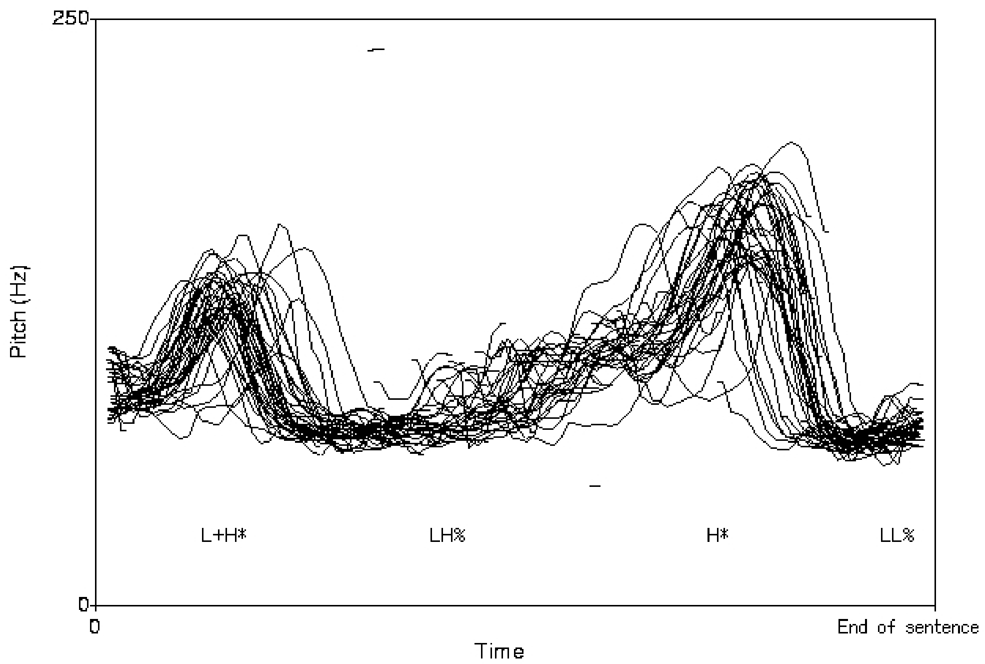

Figure 6: Pitch tracks of all responses with initial L+H* accent and late phrase boundary. The initial L+H* peak is followed by an L phrase accent and a rise to the H% boundary tone. Note the difference to figure 5, where the rise occurred much earlier. Here the rise occurs later in the sentence, and since the lengths of the phrases are more variable, the position of the rise across the sentences is more variable. As before, the later H* peak is visibly higher than the L+H* peak and is followed by an L phrase accent and L% boundary tone.

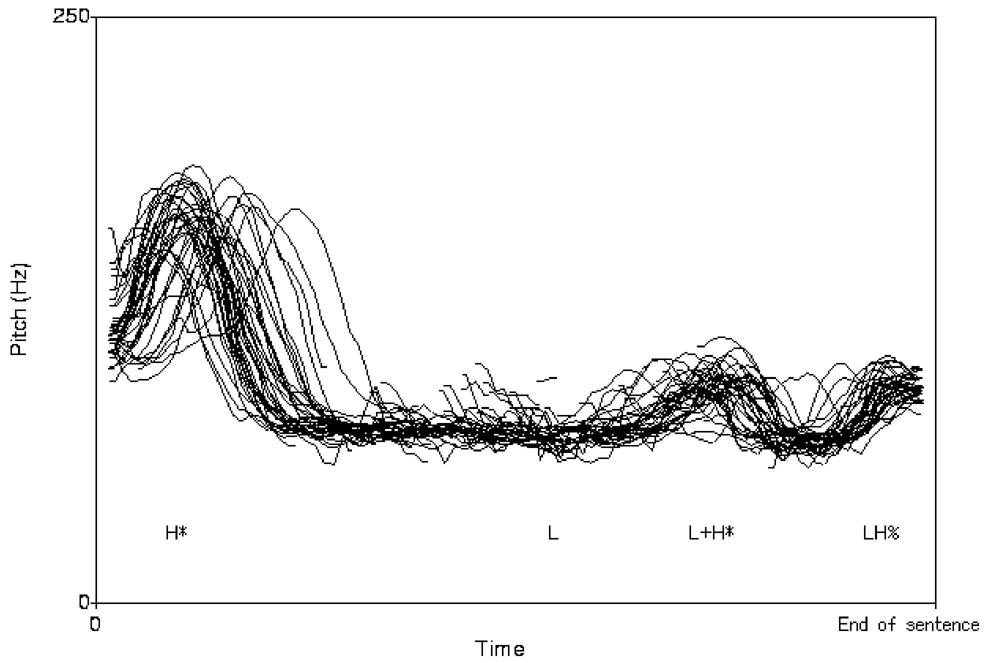

Figure 7: Pitch tracks of all responses with initial H accent and late phrase boundary. As in figure 4, the high H* peak is followed by a drop to the L phrase boundary. The sentence ends with the lower L+H* peak followed by the L phrase accent and H% boundary tone. In contrast to figures 5 and 6, it is difficult to differentiate between figures 4 and 7. Although the L phrase accent is theoretically in a different position in figures 4 and 7, in practice this cannot be detected – however, the information structure of the sentences is different and so could still affect the listeners' responses.*

For each of the 4 response types, 37 sentences were recorded by a trained male native English speaker. Recordings were made in a sound-proof recording studio, sampled at 44.8 kHz with 16 bit resolution. The volume of each recording was normalised relative to the full set, to minimise acoustic differences between recordings (other than pitch), using Normalize 0.7.7 http://normalize.nongnu.org/. Additionally, the sound files were resampled to 22.5 kHz using Voxengo r8brain http://www.voxengo.com/product/r8brain/ (to reduce disk space usage, sound file load times and to comply with requirements of experimental software, E-prime).

As this is a within-subjects design there are potential order effects which must be dealt with – such as performance deteriorating due to fatigue/boredom, performance improving due to practice, or participants adapting to the experiment e.g. working out what is being tested. To alleviate the effects of fatigue or boredom, regular breaks were inserted into the experiment, at approximately 6 minute intervals (32 trials). Although performance was not directly of interest in the experiment, it was measured, and this was later used to check for fatigue or boredom, by ensuring that participants had paid attention to the stimuli.

Finally, counterbalancing of the stimuli was used to try to account for order effects, by creating a set of different orderings to be used for different participants. A sequence of dialogues was created, consisting of eight consecutive sets of the 37 different dialogues. Within each set, the condition for a particular dialogue was pseudorandomly chosen, so that across the full sequence of dialogues each condition was represented once for each dialogue. The sequence was constrained so that the same dialogue could not occur twice in succession. Eight sequences of dialogues were then generated by repeatedly removing the first set of 37 dialogues from the front of the sequence and placing it on the end. Participants were assigned to one of these sequences.

3.3 Procedure

Before starting the experiment, participants signed a consent form, and were given instructions about the experimental procedure. It was explained that they would hear a series of mini-dialogues consisting of question and answer parts. After each dialogue they would be presented with a word, and their task was to indicate, by pressing a button, whether the word was in the previous answer. They were asked to focus on accuracy rather than speed, and told that they would be able to self-pace their responses.

Participants were provided with information about the nature of EEG recordings and asked to minimise movements during the recording, as movement artefacts would swamp signals from their brain. To minimise eye movements they were asked to fixate on a cross (+) while the audio recordings were played. They were told that there would be regular breaks where they would be able to move, and also that it was acceptable to move when the word was displayed on screen.

Participants were seated approximately 90cm in front of a monitor, in a sound-proofed testing booth. The audio recordings were played at a comfortable volume through loudspeakers positioned on either side of the screen. A PC outside the testing booth controlled presentation of the audio stimuli and the screen display, via an E-Prime program (E-Prime v1.1, Psychology Software Tools Inc., www.pstnet.com) written by the author. EEG recordings from the participant were collected by a separate PC running EEG acquisition software, Scan (Scan v4.2 and v4.3, Neuroscan, www.neuro.com), also situated outside the testing booth. At each stimulus the PC running the E-Prime program sent a signal (indicating the experimental condition) to the acquisition PC via the parallel port. This information was then used to tag the EEG recording at the correct point, thus synchronising the stimuli events with the EEG capture. (Prior to running any experiments a timing test was carried out to verify the accuracy of the synchronisation.) During the experiment, the experimenter monitored the participants via a video camera and microphone situated in the testing booth.

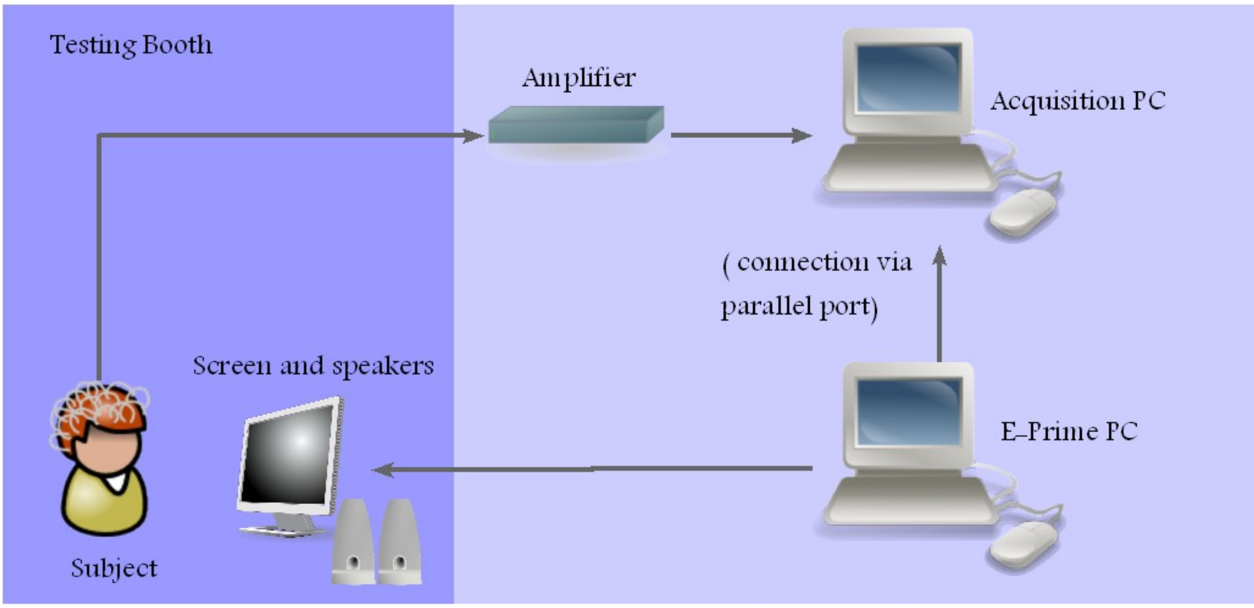

Figure 8: Experimental set-up

Dialogues were presented via speakers in blocks of 32 trials. Each dialogue was separated by a 1.5 second silent period. Within dialogues, the context and question were played, followed by a 1 second silence, followed by the response. A blue fixation cross was always displayed unless a word was being shown. The cross flashed yellow (250ms) before the questions and red (250ms) before the responses. After each response participants performed a delayed probe verification task ([50], [51]) to check that they attended to the dialogue. 1.5 seconds after the end of the response, a noun was displayed on screen (white letters on a black background). The noun either occurred in the preceding response, or did not occur in the dialogue at all. Participants had to indicate whether the noun occurred in the response or not by pressing a button. This task avoids asking participants to make judgements about the correctness or prosody of the sentence, and takes place after ERP measurements, and so should limit interference with the experimental results.

← *1.5s* →	← *(variable)* →	← *1s* →	← *(variable)* →	← *1.5s* →	← *(variable)* →
Pause	Question	Pause	Answer	Pause	Word display

Figure 9: Structure of experiment

3.4 EEG recordings

Recordings were made of the participants' EEG responses to the test sentences, measured from the onset of word of interest for both theme and rheme. Due to equipment failure, 7 recordings were made with one set of equipment (set A) and 5 with a different set (set B). Although similar, the two sets of equipment used a slightly different number of electrodes, were grounded and referenced differently, and used a different amplifier. Details of each set of equipment are given below.

For set A, the EEG was recorded from 61 silver/silver-chloride electrodes embedded in an elastic cap (QuikCap from www.neuroscan.com). The electrodes for the EEG recording were positioned according to the extended 10-20 montage [52]. The channels used were: AFZ, FCZ, CZ, CPZ, PZ, POZ, OZ, FP1, FP2, AF3, AF4, AF7, AF8, F7, F8, F5, F6, F3, F4, F1, F2, FT7, FT8, FC5, FC6, FC3, FC4, FC1, FC2, T7, T8, C5, C6, C3, C4, C1, C2, TP7, TP8, CP5, CP6, CP3, CP4, CP1, CP2, P7, P8, P5, P6, P3, P4, P1, P2, PO7, PO8, PO5, PO6, PO3, PO4, O1, O2. Two more electrodes were placed on the mastoid bones behind participants' left and right ears (M1 and M2 respectively). The recordings were referenced to the left mastoid, and later the quality of the recording was checked against the readings from the right mastoid.

Additionally electro-oculargram (EOG) recordings were made from electrodes positioned above and below the left eye (VEOGs), to detect eye blinks and vertical eye movements, and from electrodes positioned horizontally to the side of the left and right eyes respectively (HEOGs), to detect horizontal eye movements. To ensure good quality recordings, electrode impedance was kept below 5kΩ, and mastoid impedance below 2kΩ. Each channel was amplified via a Contact Precision amplifier, band-pass filtered between 0.01 and 40Hz. The EEG and EOG data was continuously sampled at 200Hz.

Equipment set B used 66 silver/silver-chloride electrodes also embedded in a QuikCap. The electrode caps had extra electrodes CB1, CB2, FC1, FC2 and FZ, with integral electrodes for M1, and M2, and did not have electrodes AF7 or AF8. The HEOG and VEOG electrodes were also integral to the cap. In this case the recordings were referenced to a central electrode rather than the left mastoid. The recordings were amplified by a SynAmps amplifier with all settings as for equipment set A.

3.5 Data Pre-processing

Before carrying out analysis of the EEG recordings the data was pre-processed using Scan 4.3 software (www.neuro.com). The initial pre-processing steps involve the removal of saturated channels, movement artefacts and eye-blinks from the data. Recordings were manually checked for saturation of channels (above or below 495μV), and where this occurred data for that period was excluded. Additionally movement artefacts were identified and excluded – typically these large artefacts occur during breaks when participants can move and stretch, so their removal has no effect on the analysis. As well as movements, eye blinks also introduce deflections in the recording, especially in channels closer to the eyes. An ocular artefact reduction procedure was run to identify blinks from the VEOG recording, and then correct for them using a regression technique [53].

The remaining pre-processing steps split the data into epochs relating to the test conditions and adjusted the waveforms so that the data can be compared across participants and trials. Each EEG recording was epoched from 100ms prior to the onset of the word of interest, to 1500ms after word onset. The waveforms for the individual epochs were smoothed over 5 points, and baselined to a common baseline determined from the 100ms before word onset. Where any channel in an epoch drifted ±50μV from the baseline, or had an amplitude more than 100μV, the epoch was rejected. 14% of trials were rejected due to baseline drift, probably due to the length of the epoch used.

Average ERP waveforms were produced for all of the remaining processed epochs for each participant, averaged across each condition (16 conditions: word position(first/second), pitch accent(H*/L+H*), expected information structure (theme/rheme), boundary position (early/late)). Finally grand average waveforms were produced, showing the average across all participants for each condition.

4 Results

4.1 Behavioural results

All subjects performed well in the task, as expected, given its straightforward nature (mean score 98% ± 1%, chance 50%). The low number of incorrect responses indicates that the subjects attended to the dialogues, so no subjects were excluded on the basis of poor performance, and incorrect trials were not excluded.

4.2 ERP results

Due to time constraints, analyses were carried out for only a subset of cases, the conditions involving the first word in the sentence. The first word was chosen as the effects of the timing of the phrase boundary would be reduced or absent, important since the correct stimulus point to use is unclear, since word onset, peak pitch accent and phrase boundary position were all identified as possible trigger points. In the case of the first word, the word onset is a reasonable approximation to the position of the pitch accent and most effects of interest should occur before the intonational phrase boundary is reached.

Grand average waveforms are shown for these conditions in figure 10 (early phrase boundaries) and figure 11 (late phrase boundaries) for the 59 scalp electrodes common to both sets of equipment. The most prominent feature revealed by visual inspection of the waveforms is a relatively late difference in the waveforms between conditions where different questions were used (apparent for early phrase boundaries only) i.e. where the expected information structure is different. This is most noticeable in parietal locations, for example at electrode CPZ shown in figures 12 and 13. Additionally there seem to be differences between appropriate and inappropriate intonations in frontal locations in a similar time-frame, for example at electrode FCZ shown in figure 14 and 15.

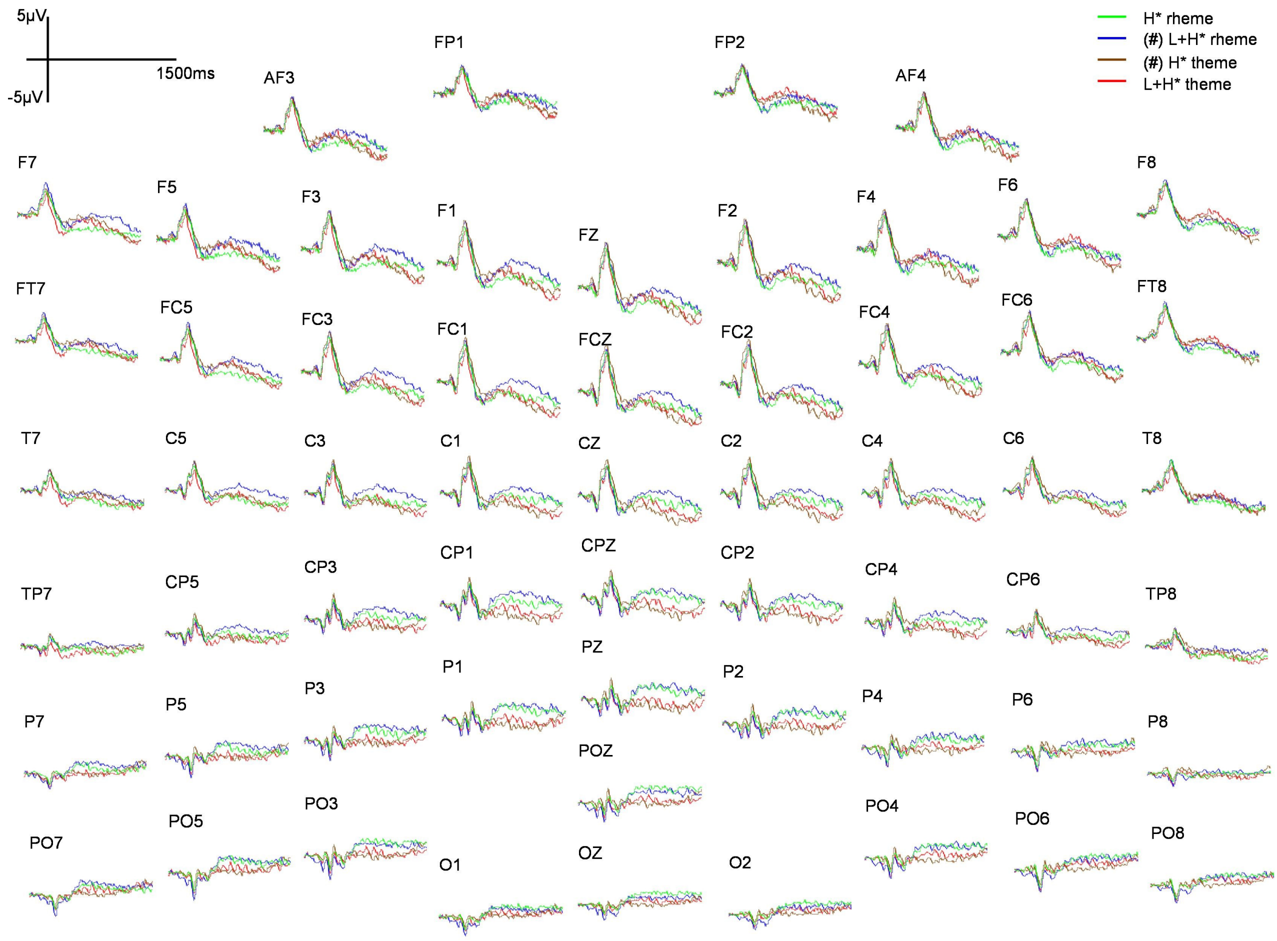

Figure 10: Grand average ERP waveforms (11 subjects) elicited at the 59 scalp electrodes in response to the first accented word in a sentence with an early phrase boundary. Words were either accented with H or L+H* pitch accents and the word was expected to be either a theme or a rheme. Waveforms are identified by pitch accent and expected information structure. Inappropriate intonation is marked by (#). ERPs are shown over an epoch of 1500ms, starting from the acoustic onset of the target word at 0ms. Mean number of trials (±standard deviation) contributing to the ERPs were 34.64 ±2.34 for the H* with rheme condition, 34.09±3.11 for the L+H* with rheme condition, 34.91±2.39 for the L+H* with theme condition and 34.18±3.09 for the H* with theme condition. Positivity is plotted upwards.*

AF = anterior-frontal, FP = fronto-parietal, FC = fronto-central, C= central, T = temporal, CP = centro-parietal, TP = tempero-parietal, FT = fronto-temporal, P= parietal, PO = posterior, O = occipital

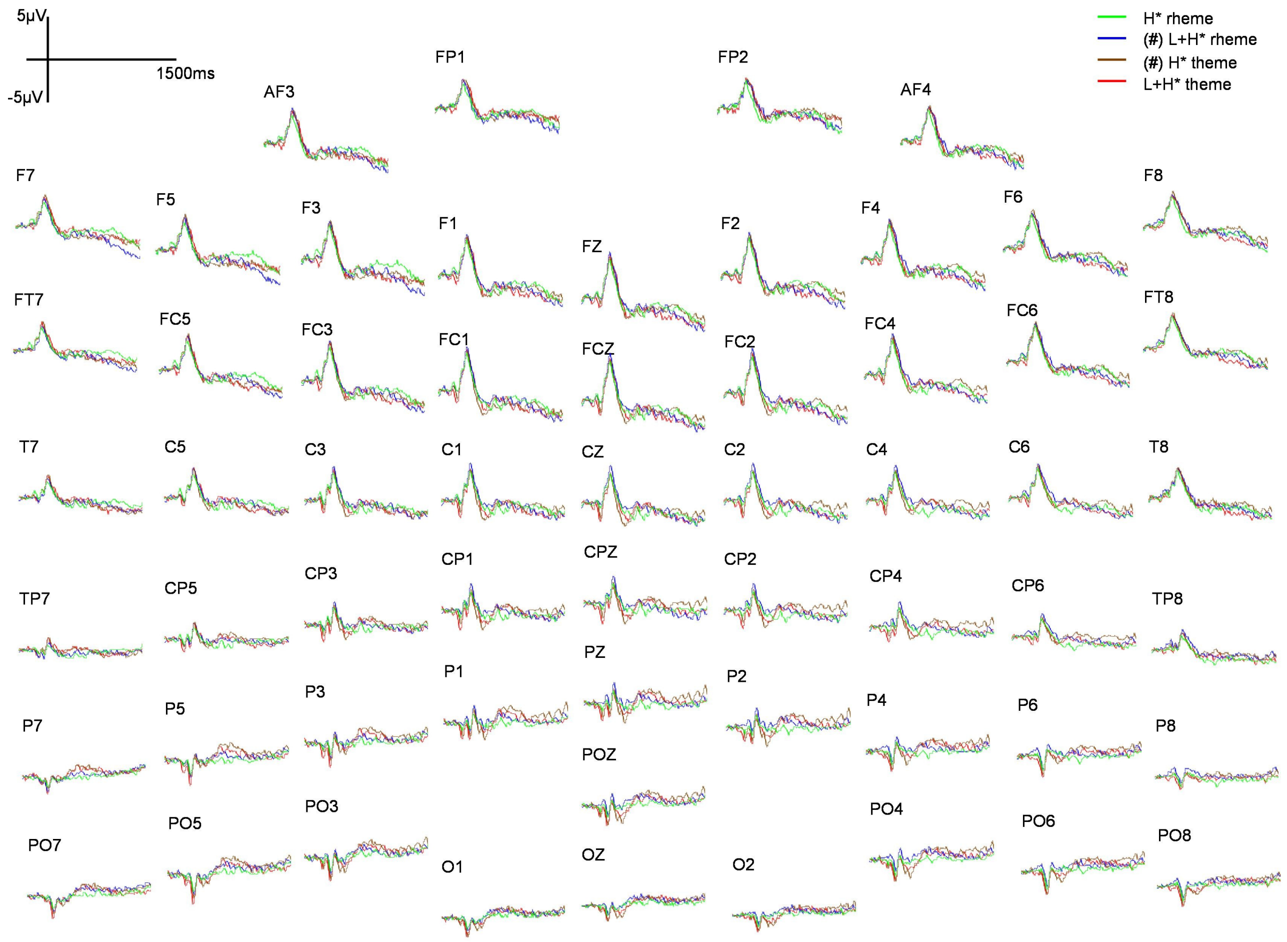

Figure 11: Grand average ERP waveforms (11 subjects) elicited at the 59 scalp electrodes in response to the first accented word in a sentence with a late phrase boundary. Words were either accented with H or L+H* pitch accents and the word was expected to be either a theme or a rheme. Waveforms are identified by pitch accent and expected information structure. Inappropriate intonation is marked by (#). ERPs are shown over an epoch of 1500ms, starting from the acoustic onset of the target word at 0ms. Mean number of trials (±standard deviation) contributing to the ERPs were 34.27 ±2 for the H* with rheme condition, 34.18 ±2.6 for the L+H* with rheme condition, 34.18 ±3.68 for the L+H* with theme condition and 34.55 ±2.98 for the H* with theme condition. Positivity is plotted upwards.*

AF = anterior-frontal, FP = fronto-parietal, FC = fronto-central, C= central, T = temporal, CP = centro-parietal, TP = tempero-parietal, FT = fronto-temporal, P= parietal, PO = posterior, O = occipital

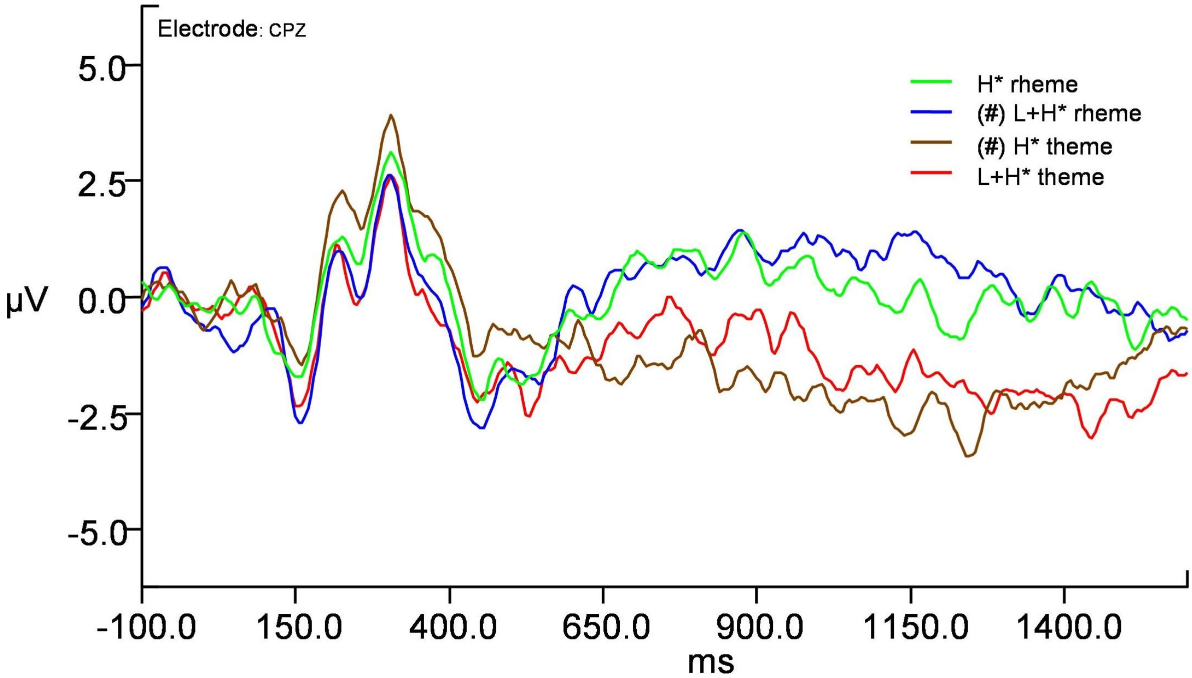

Figure 12: Grand average ERP waveforms (11 subjects) elicited at the CPZ scalp electrodes in response to the first accented word in a sentence with an early phrase boundary. Words were either accented with H or L+H* pitch accents and the word was expected to be either a theme or a rheme. Waveforms are identified by pitch accent and expected information structure. Inappropriate intonation is marked by (#). ERPs are shown over an epoch of 1500ms, starting from the acoustic onset of the target word at 0ms.*

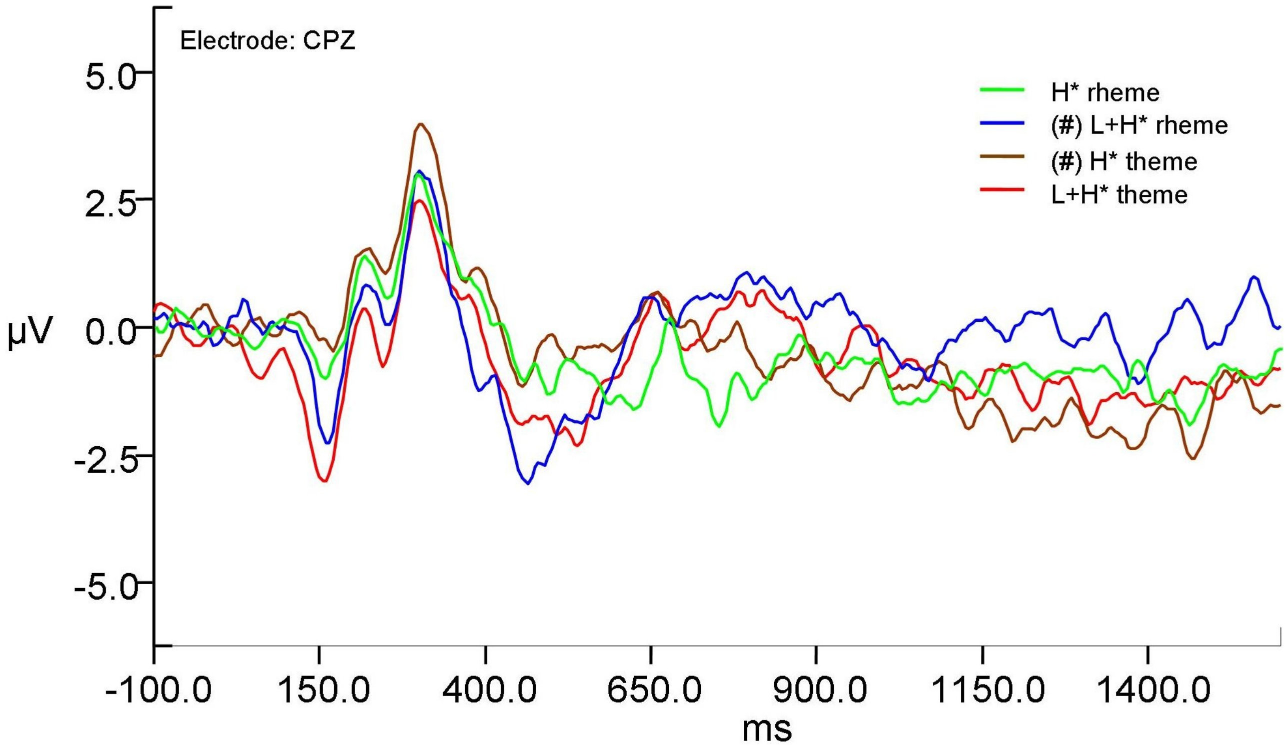

Figure 13: Grand average ERP waveforms (11 subjects) elicited at the CPZ scalp electrode in response to the first accented word in a sentence with a late phrase boundary. Words were either accented with H or L+H* pitch accents and the word was expected to be either a theme or a rheme. Waveforms are identified by pitch accent and expected information structure. Inappropriate intonation is marked by (#). ERPs are shown over an epoch of 1500ms, starting from the acoustic onset of the target word at 0ms.*

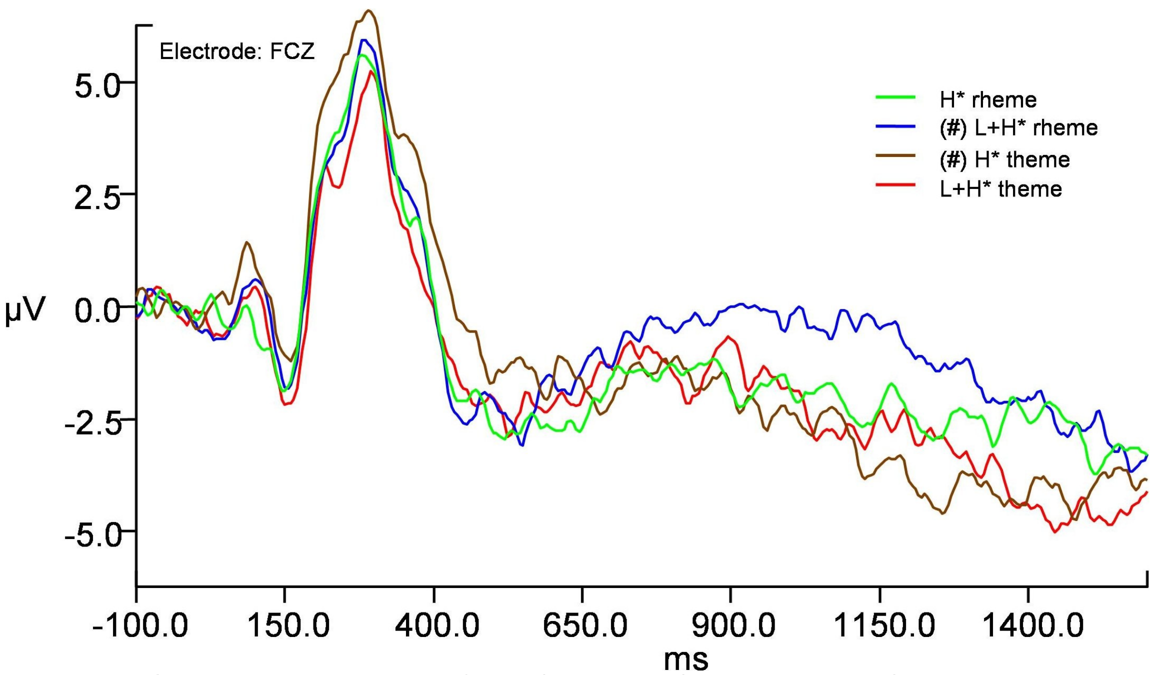

Figure 14: Grand average ERP waveforms (11 subjects) elicited at the FCZ scalp electrode in response to the first accented word in a sentence with a late phrase boundary. Words were either accented with H or L+H* pitch accents and the word was expected to be either a theme or a rheme. Waveforms are identified by pitch accent and expected information structure. Inappropriate intonation is marked by (#). ERPs are shown over an epoch of 1500ms, starting from the acoustic onset of the target word at 0ms.*

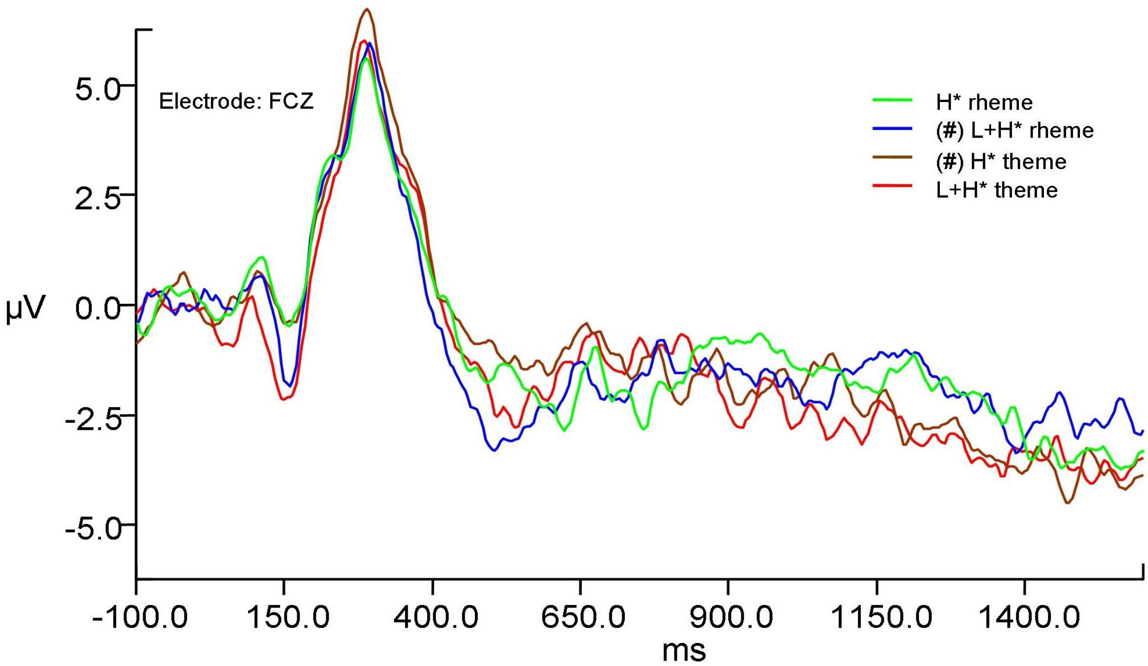

Figure 15: Grand average ERP waveforms (11 subjects) elicited at the FCZ scalp electrode in response to the first accented word in a sentence with a late phrase boundary. Words were either accented with H or L+H* pitch accents and the word was expected to be either a theme or a rheme. Waveforms are identified by pitch accent and expected information structure. Inappropriate intonation is marked by (#). ERPs are shown over an epoch of 1500ms, starting from the acoustic onset of the target word at 0ms.*

4.3 Data Analysis

All analysis was carried out on the mean voltage amplitude relative to a 100ms pre-stimulus baseline. Time windows used were 200-300ms, 300-400ms, 400-500ms (around the typical windows for P300 and N400 components, both effects which have been found in related research), and 800-1100ms, 1100-1300ms, 1300-1500ms (based on inspection of the waveforms). Locations for an initial analysis were anterior (FC3, FC1, FCZ, FC2, FC4, F3, F1, FZ, F2, F4) and posterior (CP3, CP1, CPZ, CP2, CP4, P3, P1, PZ, P2, P4), based on the known locations of N400 and P300 effects and inspection of the waveforms.

The data was submitted to a repeated measures Analysis of Variance (ANOVA) using the Greenhouse-Geisser correction where appropriate (to correct for factors whose variances violate the sphericity assumption of ANOVA), as determined by Mauchly's test. The within-subjects factors for the ANOVA were pitch accent (H* / L+H*), expected information structure (theme/rheme), phrase boundary (early/late), time window and location. Appropriateness of intonation is represented by the interaction of the pitch accent and expected information structure factors.

No effect of intonation appropriateness was found in any time window, but separate effects of pitch accent and expected information structure were apparent. Effects of pitch accent were almost entirely restricted to the early time windows while effects of expected information structure were present in all time windows. A main effect of pitch accent was present in the 300-400ms and 400-500ms time windows ($p=0.039$ $F(1,10)=5.657$ and $p=0.029$ $F(1,10)=5.659$ respectively), and an interaction effect of pitch accent and phrase boundary was present in the 200-300ms time window ($p=0.016$, $F(1,10)=8.477$) indicating early effects of pitch accent which differ for the different positions of the phrase boundary. Additionally interaction effects of location and expected information structure were found in the 200-300ms and 300-400ms time windows ($p=0.02$, $F(1,10)=7.62$ and $p=0.039$, $F(1,10)=5.657$ respectively). Finally, a main effect of expected information structure was present in the 400-500ms time window ($p=0.039$, $F(1,10)=5.659$).

Considering the later time windows, a main effect of expected information structure was found in the 1100-1300ms and 1300-1500ms time windows ($p=0.019$, $F(1,10)=7.838$, and $p=0.032$, $F(1,10)=6.158$ respectively), and interaction effects of expected information structure and phrase boundary in the 800-1100ms, 1100-1300ms and 1300-1500ms time window ($p=0.019$, $F(1,10)=7.794$; $p=0.042$, $F(1,10)=5.411$

and p=0.037, F(1,10)=5.811), indicating that the effects of expected information structure differ for different positions of the phrase boundary. Finally, a late main effect of accent was present in the 1100-1300ms time window (p=0.017, F(1,10)=8.229). This is of interest since this time window was identified as having a possible effect of intonation appropriateness. Further analyses were carried out to characterise the effects of pitch accent and expected information structure, and the interactions with phrase boundary position.

4.3.1 Effects of expected information structure

Expected information structure relates to whether the first word is expected to be a theme or a rheme, and is determined by the context question. e.g.

Question and answer pair	Expected theme/rheme on first word?
Q: *I know EMMA will marry ARNIE. But who will marry MANNY?* *A: (ANNA) (will marry MANNY).* *H* L L+H* LH%*	Rheme
Q: I know EMMA will marry ARNIE. But what about ANNA? *A: (ANNA) (will marry MANNY).* *L+H* LH% H* LL%*	Theme
Q: *I know EMMA will marry ARNIE. But who will marry MANNY?* *#A: (ANNA) (will marry MANNY).* *L+H* LH% H* LL%*	Rheme
Q: I know EMMA will marry ARNIE. But what about ANNA? *#A: (ANNA) (will marry MANNY).* *H* L L+H* LH%*	Theme

The results above indicate that there is an effect of expected information structure over the 800-1300ms time period, but that it is different for different phrase boundaries. Figure 16 below shows the mean voltages over the three time windows for when the first word is expected to be a theme versus a rheme, contrasted between sentences which have an early phrase boundary and those which have a late phrase boundary. Where the sentence has an early phrase boundary, the mean voltages vary depending on whether the first word is expected to be a theme or a rheme, with the voltages for an expected rheme consistently more negative than for an expected theme. Where the sentence has a late phrase boundary, this effect is not present. A repeated measures ANOVA as before, but including data only from the late phrase boundary condition, confirmed that there were no significant effects of expected information structure in any of the 3 time windows (p>0.4, F(1,10)<0.7 in all cases).

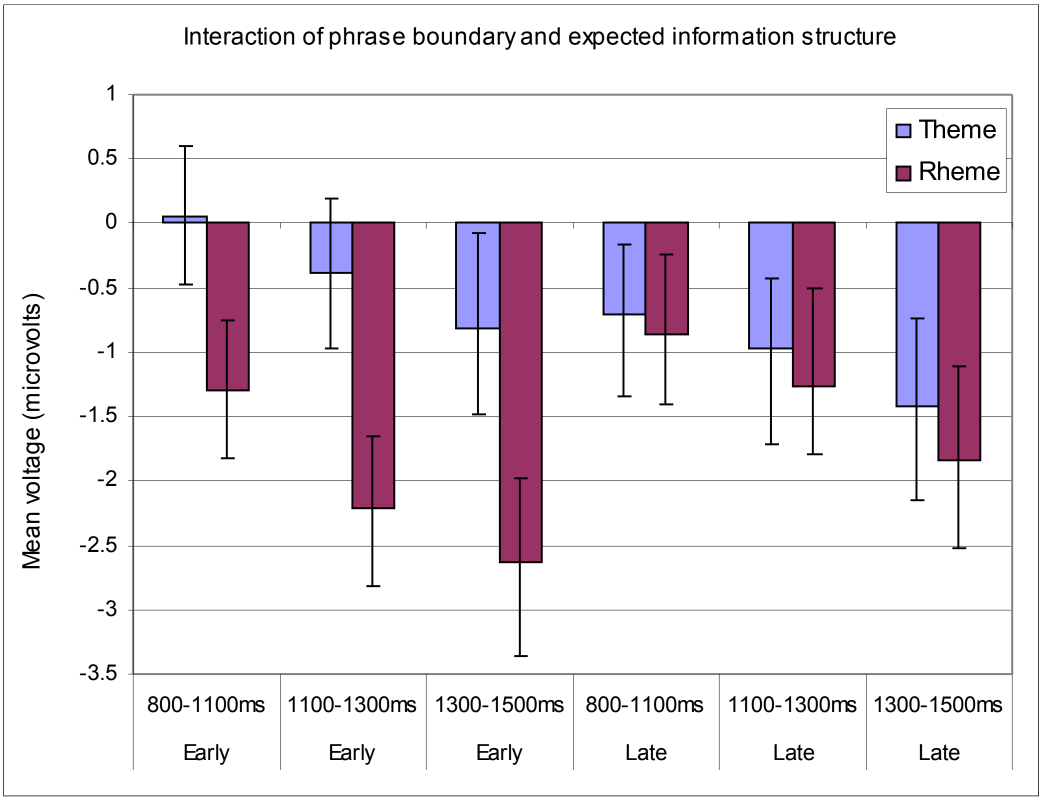

Figure 16: Effects of expected information structure in relation to phrase boundary, over time

For the early phrase boundary, the expected information structure has a pronounced effect, with the difference between theme and rheme manifesting itself as a sustained centro-parietal negativity 600-1500ms from onset of the first word in the sentence, peaking at approximately 1100ms after onset (figure 17).

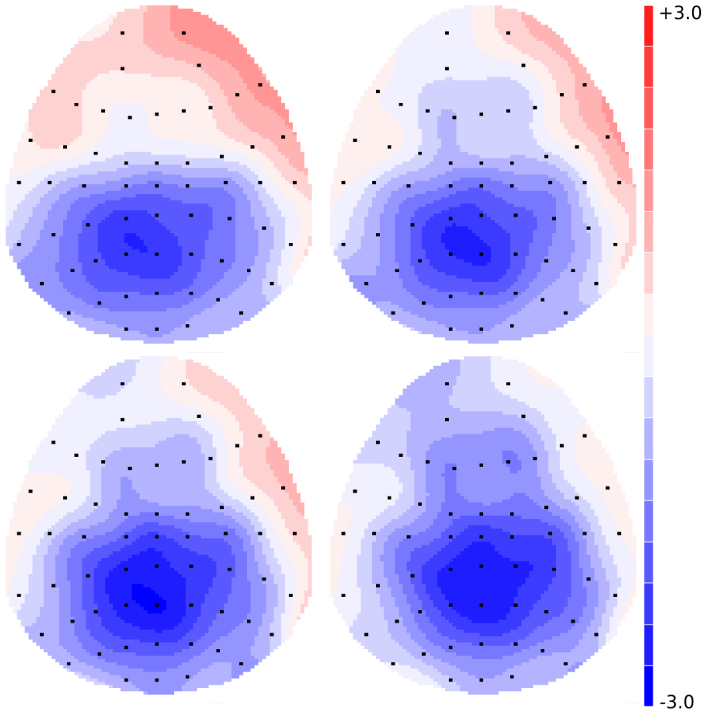

Figure 17: Topographic distribution of the effect of expected information structure, calculated by subtracting the mean voltages at each electrode for rheme from those for theme. Voltages are in microvolts between -3μV and 3μV according to the coloured scale. Each head-shot corresponds to an average over 200ms: top left 600-800ms, top right 800-1000 ms, below left 1000-1200ms, below right 1200-1400ms.

The topographic distribution of the effect (figure 17) resembles that of the N400 and so a further test was carried out to assess its temporal and spatial distribution. A repeated measures ANOVA was carried out over the 3 time windows covering 800-1500ms, and over the CP set of electrodes spanning the centro-parietal area (TP7, CP5, CP3, CP1, CP2, CP4, CP6, TP8, omitting CPZ in order to analyse hemisphere effects). Within-subjects factors were pitch accent, expected information structure, time window, and location (superior, superior middle, inferior middle, inferior) and hemisphere (left/right). The analysis revealed a main effect of expected information structure ($p=0.001$, $F(1,10)=24.843$), confirming that the mean voltages over the full 800-1500ms time window for expected theme are lower than expected rheme in this location (-1.3μV and 0.39μV respectively). Furthermore there were interaction effects of location and expected information structure ($p=0.001$, $F(1.237,12.371)=18.338$) indicating that the effect varies across the scalp (Figure 18). Lastly there was an interaction effect of expected information structure, hemisphere and time window ($p=0.018$, $F(2,20)=4.917$) indicating changes in the hemispheric distribution over time. From figure 18 the differences in hemisphere relate to a slight bias of the topographic distribution to the left. Overall the distribution is similar to the N400 distribution.

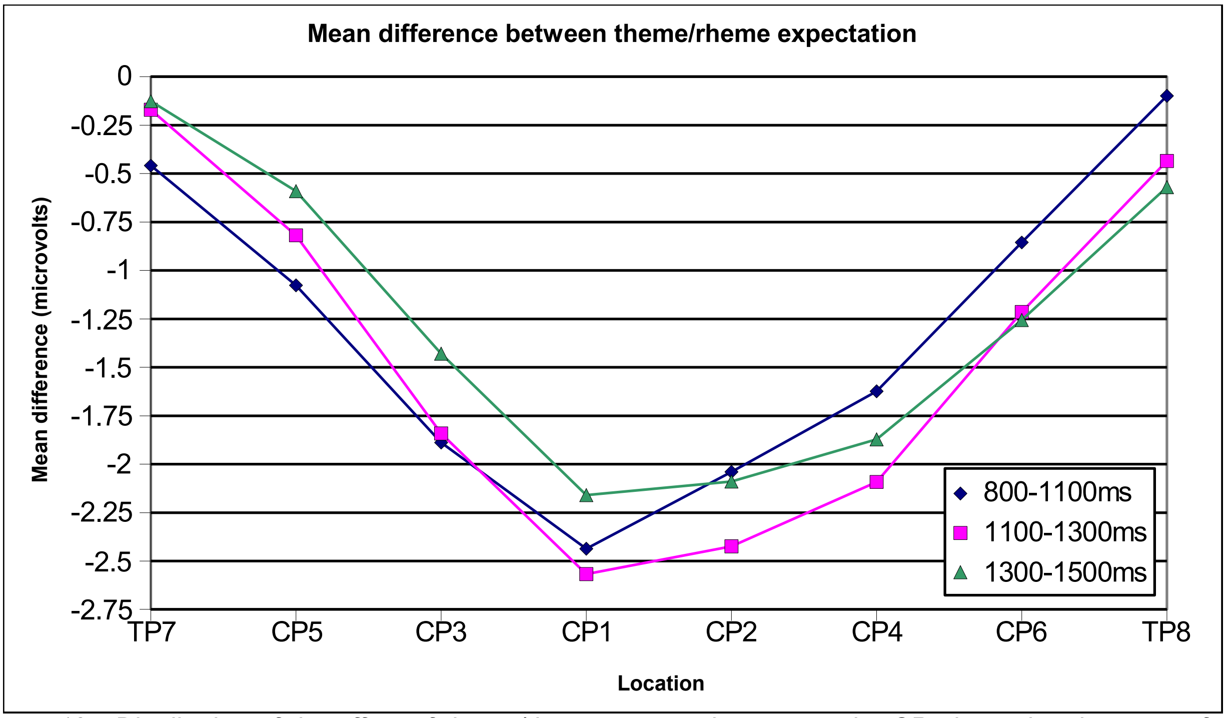

Figure 18: Distribution of the effect of theme/rheme expectation across the CP electrodes, in terms of the mean difference in theme and rheme expectation, over time

4.3.2 Effects of pitch accent

Pitch accent relates to whether the first word is accented with an H* or L+H* accent (regardless of the expected information structure), i.e.

Question and answer pair	Pitch accent?
Q: *I know EMMA will marry ARNIE. But who will marry MANNY?* *A: (ANNA) (will marry MANNY).* *H* L L+H* LH%*	H*
Q: I know EMMA will marry ARNIE. But what about ANNA? *A: (ANNA) (will marry MANNY).* *L+H* LH% H* LL%*	L+H*
Q: *I know EMMA will marry ARNIE. But who will marry MANNY?* *#A: (ANNA) (will marry MANNY).* *L+H* LH% H* LL%*	L+H*
Q: I know EMMA will marry ARNIE. But what about ANNA? *#A: (ANNA) (will marry MANNY).* *H* L L+H* LH%*	H*

The initial analysis revealed effects of pitch accent in the 200-500ms after onset of the first word, specifically an interaction between pitch accent and phrase boundary position in the 200-300ms window and a main effect of pitch accent in the 300-400ms and 400-500ms windows.

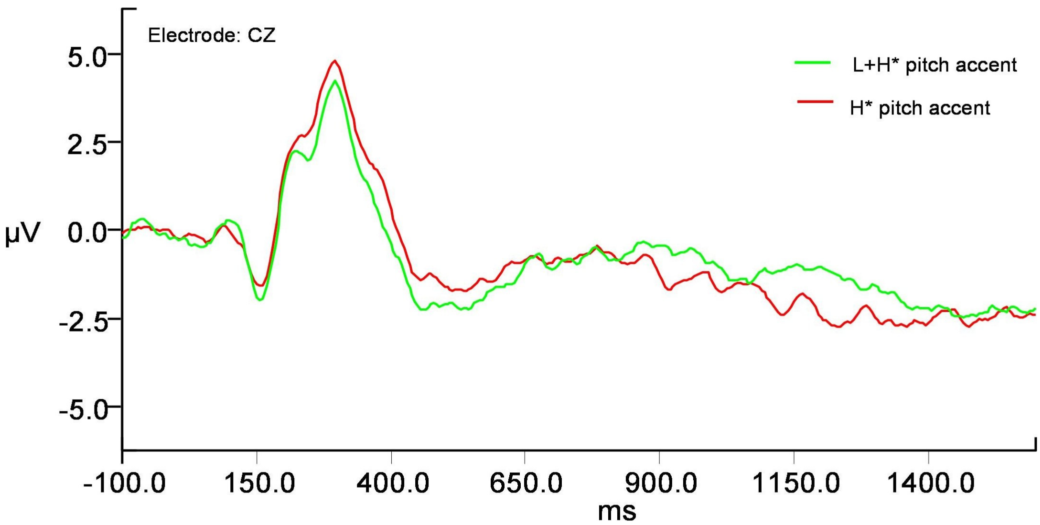

Figure 19: Grand average ERP waveforms (11 subjects) elicited at the CZ scalp electrode in response to the first accented word in a sentence (averaged over both types of phrase boundary). Words were either accented with H (red) or L+H* (green) pitch accents . ERPs are shown over an epoch of 1500ms, starting from the acoustic onset of the target word at 0ms.*

Comparisons of the mean voltage amplitudes for different levels of pitch accent in the first time window indicate that the pitch accent has no effect in the 200-300ms time window for the late phrase boundary (confirmed by ANOVA on accent, location, and site as before, for the late phrase boundary only, p=0.95, F(1,10)=.004). This is a slightly odd result as there is no reason for the effects of pitch accent to be different between phrase boundary positions at this point. For this reason the remainder of the analysis on pitch accent was carried out on the 300-500ms time period, and across both phrase boundary position cases.

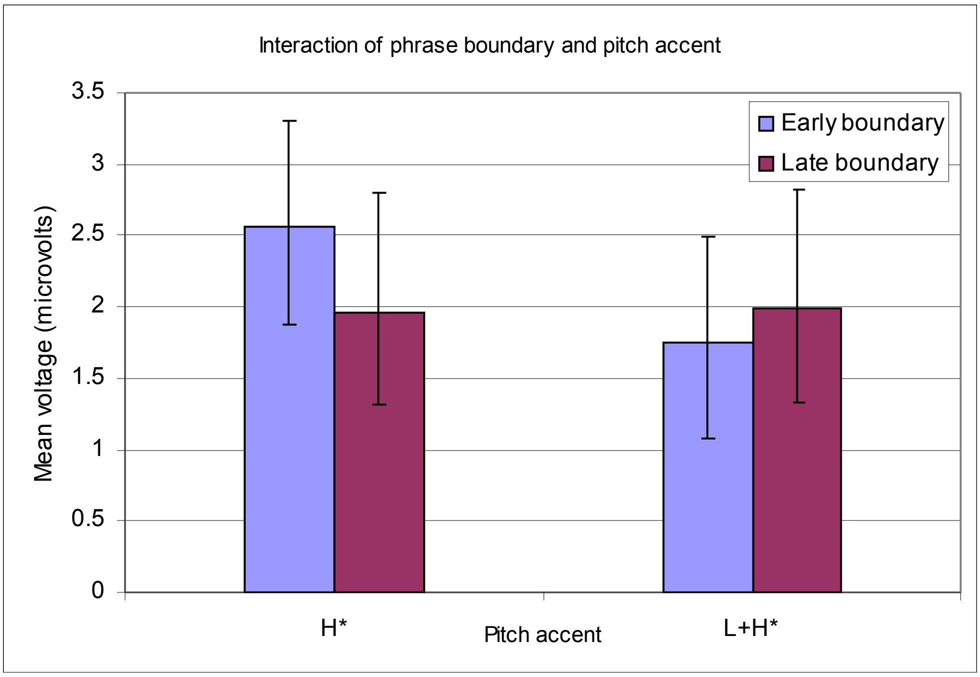

Figure 20: Comparison of the effect of pitch accent on the mean voltage, by phrase boundary position, over the 200-300ms time window. There is an effect of pitch accent in sentences with an early boundary but not in sentences with a late boundary.

During the 200-400ms after word onset, the effect of pitch accent is broadly distributed across the scalp, mostly symmetrical across the midline (figure 21). The effect is quite small, peaking at around -1.2µV. In an effort to characterise the distribution of the effect, a repeated measures ANOVA was carried out over the 200-400ms time window over two sets of electrodes, superior (FC1, FCZ, FC2, C2, CP2, CPZ, CP1, C1) and inferior (F3, FZ, F4, C4, P4, PZ, P3, C3, F3). Within-subjects factors were pitch accent, time window, and location (superior, inferior). An interaction effect of location and pitch accent would indicate that the effect is stronger in superior areas relative to inferior areas. However no significant interaction effect of location and pitch accent was found (p=0.248, F(1,10)=1.504). The same ANOVA with a wider spread of electrodes across the scalp (superior (FC1, FCZ, FC2, C2, CP2, CPZ, CP1, C1) and inferior (F5, FZ, F6, C6, P6, PZ, P5, C5, F5)) also showed no interaction effect of pitch accent and location (p=0.248, F(1,10)=1.504). Finally, the same ANOVA with electrodes positioned across the scalp left to right across the centre (C1, C3, C5, T7, C2, C4, C6, T8) with an additional factor of hemisphere (left/right) also yielded no significant effect ($p=0.214$, $F(1.122, 11.217)=1.755$).

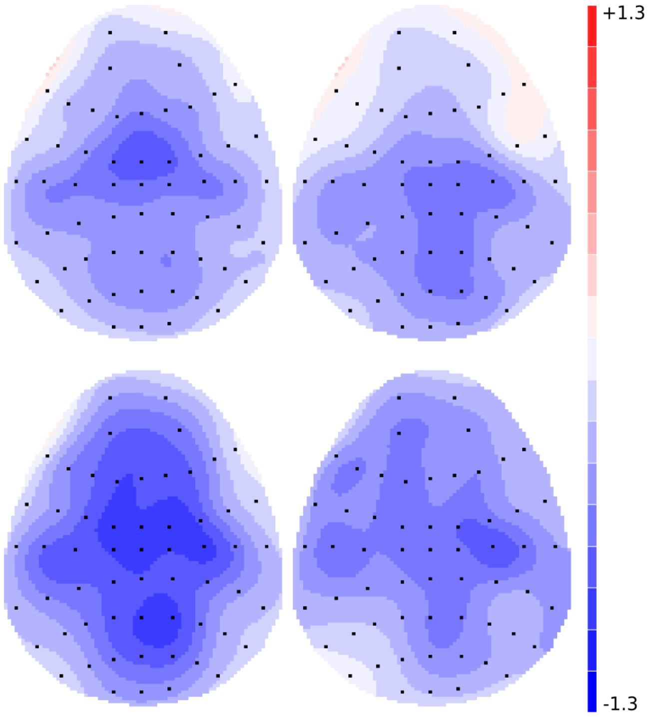

Figure 21: Topographic distribution of the effect of pitch accent, calculated by subtracting the mean voltages at each electrode for H from those for L+H*. Voltages are between -1.3µV and 1.3µV according to the coloured scale. Each head-shot corresponds to an average over 50ms: top left 200-250ms, top right 250-300 ms, below left 300-350ms, below right 350-400ms.*

4.4 Effects of intonation appropriateness

Intonation appropriateness relates to whether the pitch accent on the first word corresponds to the expected information structure,and is determined by both the actual pitch accent on the first word and the expected information structure of the sentence i.e.

Question and answer pair	Intonation appropriateness?
Q: *I know EMMA will marry ARNIE. But who will marry MANNY?* *A: (ANNA) (will marry MANNY).* *H* L L+H* LH%*	Appropriate
Q: I know EMMA will marry ARNIE. But what about ANNA? *A: (ANNA) (will marry MANNY).* *L+H* LH% H* LL%*	Appropriate
Q: *I know EMMA will marry ARNIE. But who will marry MANNY?* *#A: (ANNA) (will marry MANNY).* *L+H* LH% H* LL%*	Inappropriate
Q: I know EMMA will marry ARNIE. But what about ANNA? *#A: (ANNA) (will marry MANNY).* *H* L L+H* LH%*	Inappropriate

The initial analysis showed that there were two time windows (400-500ms and 1100-1300ms) where main effects of both expected information structure and pitch accent occurred. The waveforms over these time periods both share the characteristic that the intonationally inappropriate waveforms are between the intonationally inappropriate waveforms (figure 22).

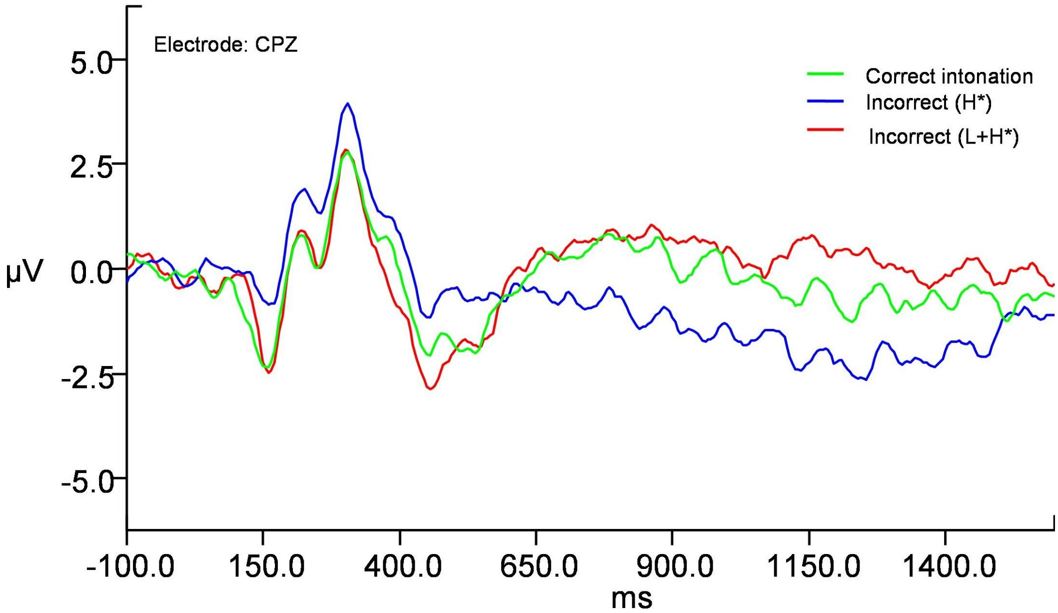

Figure 22: Grand average ERP waveforms (11 subjects) elicited at the CPZ scalp electrode in response to the first accented word in a sentence (averaged over early and late phrase boundaries). Intonation is either correct (green), incorrectly accented with H (blue) or incorrectly accented with L+H* (red). ERPs are shown over an epoch of 1500ms, starting from the acoustic onset of the target word at 0ms.*

In this configuration the interaction effect of pitch accent and expected information structure could fail to show an effect of intonation appropriateness. Therefore a new factor ("intonation") was introduced to the model, with 4 levels, one for each of the possible combinations of pitch accent and expected information structure. Rerunning the ANOVA with this factor gave significant effects of intonation in both time windows (400-500ms: $p=0.025$, $F(3,30)=3.574$; 1100-1300ms: $p=0.002$, $F(3,30)=6.097$). No interaction effect of location and intonation was found (400-500ms: $p=0.13$, $F(3,30)=2.033$; 1100-1300ms: $p=0.363$, $F(3,30)=3.63$), and no further attempt was made to measure the topographic distribution as the differences in voltage are very small.

Comparisons of the mean scalp voltages for inappropriate and appropriate intonation are shown for the 400-500ms time window in figure 23 and for 1100-1300ms in figure 24. In both time windows the two appropriate intonations have the same mean voltage and appear to be similar. In the 400-500ms time window one inappropriate intonation (theme with H* pitch accent) is slightly higher than the appropriate cases, while the other inappropriate intonation (rheme with L+H* pitch accent) is slightly lower. The polarity of the differences is reversed in the 1100-1300ms time window. It is most likely that the effect seen in the 1100-1300ms time window is the same effect as in the 400-500ms window, just on the second accented word in the sentence.

Since the second word always has the opposite pitch accent to the first word, the effect in relation to the pitch accent on the first word is reversed. The only significant difference in mean voltage which could be detected was the difference between the two inappropriate intonations (400-500ms: p=0.019, F(1,10)=7.837; 1100-1300ms: p=0.003, F(1,10)=15.771).

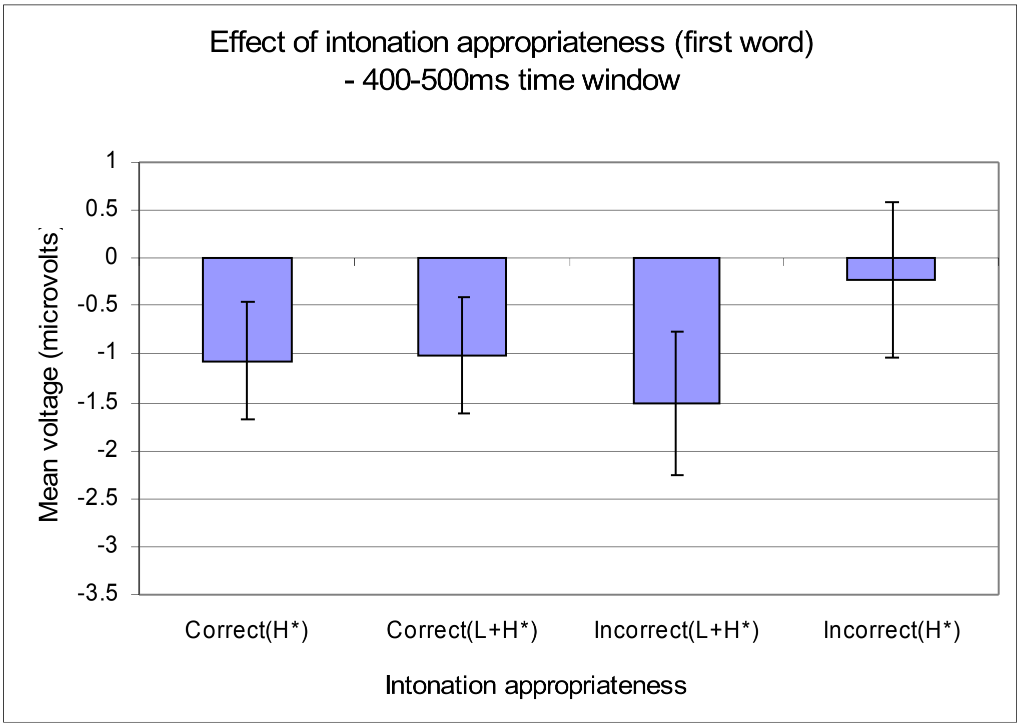

*Figure 23: Mean voltages for appropriate and inappropriate intonation 400-500ms after onset of the accented word. The actual pitch accent on the **first** word is given in brackets.*

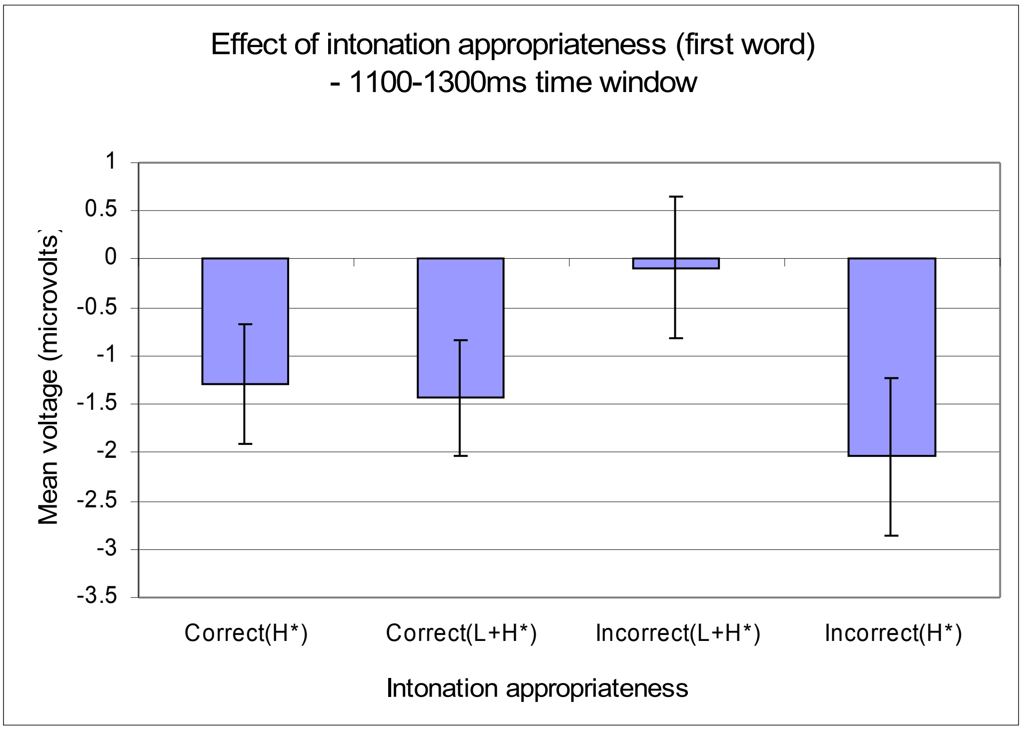

*Figure 24: Mean voltages for appropriate and inappropriate intonation 1100-1300 ms after onset of the accented word. The actual pitch accent on the **first** word is given in brackets.*

5 Discussion

The data analysis identified three effects: a late centro-parietal effect of expected information structure, an earlier broadly-distributed effect of pitch accent and a small effect of intonation appropriateness.

5.1 Intonation

The goal of this study was to investigate the neural response to inappropriate intonation in the context of Steedman's theory [1] relating information structure to intonational tunes. The theory states that the H* LL% tune marks rhemes and the L+H* LH% tune marks themes, and so predicts that switching the tunes on themes and rhemes will elicit a neural response during sentence processing of either theme or rheme. In particular, the theory predicts that the response will be during semantic processing. The experiment was designed to test the theory by comparing the responses to sentences with rhemes and themes marked with the corresponding tune or the alternative tune.

The data analysis shows a small but significant effect of intonation appropriateness. The data suggests that approximately 400-500ms after word onset, an inappropriate H* LL% tune marking a rheme has a positive effect on mean scalp voltage relative to appropriate intonation with either tune. Similarly an inappropriate L+H* LH% tune marking a rheme has a negative effect relative to appropriate intonations. The difference in voltage is very small, under 1μV, and the only significant difference in voltage was between the two inappropriate tunes. It is therefore possible that the effect of one of these tunes is not significantly different from the appropriate tunes. However given the number of subjects and the size of the effect it is also possible that the effect would be significant under more subjects. As such the data supports the prediction that an effect of inappropriate intonation affects processing of both themes and rhemes.

Additionally, the latency of the effect, 400-500ms after word onset, is close to the latency of the N400 effect, which is known to relate to semantic processing, and whose amplitude corresponds to the difficulty of semantic integration. It was not possible to quantify the topographic distribution of the effect, again most probably because of the small number of subjects, small voltage differences involved and the corresponding lack of power in the tests. Without comparing the effect distribution with the typical N400 distribution it is difficult to categorise the intonation appropriateness effect as an N400 effect, although this seems to be a strong possibility. (Note that if the effect could be classified as an N400, and the L+H* LH% and H* LL% tunes have opposite effects, the correspondence of the magnitude of the N400 effect to semantic integration difficulty would be contradicted.)

The effect observed in the 1100-1300ms time window is probably the same intonation appropriateness effect occurring for the second word in the sentences. Analysing based on the onset of the first word in the sentence does not give particularly good time-locking, which may account for the difference in magnitude and duration of the effect, although it seems likely that at the second word the effect is at least larger in magnitude, if not latency. In terms of the relationship between intonation and information structure this makes sense: at the end of the sentence the possible completions are fully defined, so an anomalous accent should have a larger effect, as in some sense it is more anomalous than if it had appeared earlier in the sentence.

Previous studies have investigated the link between pitch accent and information structure, but have not shown the effects found here. In other studies either no effect or a P300 effect on an inappropriately intoned first word was shown, and an N400 effect on an inappropriately intoned second word. The present study shows the same effect on both first and second words, most closely related to an N400 There are several possible reasons that the results do not agree. Firstly, most studies use a prosodic judgement task [44],[46],[47], where subjects make a decision on the prosodic appropriateness of the sentences as they listen to them. This typically elicits a task-related P300 effect relative to the onset of the first target word in the sentence e.g. [44]. A P300 could mask the small negative effect of intonation appropriateness at 400ms. In this study the task was selected to avoid P300 effects. Secondly, other studies have compared appropriate and inappropriate application of the H* pitch accent in relation to a missing pitch accent, but not in relation to the L+H* pitch accent. Since the effect of intonation appropriateness seems to polarise L+H* and H* accents, the effect will probably be largest when comparing the two pitch accents, and smaller when comparing H* with a missing accent. The magnitude of the effect is already small, and so comparing L+H* and H* increases the likelihood of finding a significant effect.

Further work is needed to quantify the effect of intonation appropriateness on the second word, although the analysis suggests that the effect is the same or larger than that on the first word. Additionally it will be important to investigate the correct stimulus for the effect as this could have implications for the theory linking information structure and intonation, specifically relating to whether the phrase boundary is an important determiner of theme or rheme. Potential candidates for the stimulus include word onset, pitch accent peak, and phrase boundary (in the case of the second word). This is a matter of baselining the data already collected against the different time points for these features.

The polarised L+H* / H* response is interesting, as it may provide new insight into the debate over the definition of L+H* and H*. For example, it would be possible to investigate the effects of changing different characteristics of L+H* and H* to directly determine which characteristics make an L+H* more like an H* and vice versa.

Finally, it will be necessary to perform the experiment with more subjects in order to increase the power of the statistical tests, to identify the effect of intonation appropriateness more clearly.

5.2 Expected information structure

The effect of expected information structure was already identified by Hruska et al [47] and Johnson et al [46], and both authors identify the effect as a likely closure positive shift (rheme has a positive effect in relation to no pitch accent; in the present analysis the effect of rheme was subtracted, hence the negative effect shown). The analysis here shows the effect is consistent with the CPS. The CPS is defined as a positive shift occurring in response to intonational phrase boundaries, and is broadly distributed and symmetrica, and in this case would reflect the effort of integrating the rheme as the answer to the previous question. The time windows analysed only included the early phrase boundary and not the late phrase boundary, however the effect only occurred in the early phrase boundary case, supporting the view that the effect is a CPS. Further work is necessary to confirm that the effect also occurs in conjunction with the late phrase boundary.

The topographical distribution of the effect reported here does differ slightly from the typical CPS reported in the literature [42]. The effect in the present study is concentrated on the centro-parietal region, although also symmetric, and resembles an N400 effect in terms of its distribution. It is possible that the difference is related to the use of an accented theme rather than an unaccented theme in Hruska's and Johnson's experiments, and this could be investigated further.

5.3 Pitch accent

An effect of pitch accent was observed in the early time windows (200-500ms) after onset of the first accented word. The effect was rather small and with so few subjects, difficult to characterise topographically, although it appears to have a broad symmetrical distribution peaking towards the midline of the scalp. Given the early timescale and the differences between L+H* and H*, one possible interpretation of the effect is that it simply relates to the detection of different pitches. Neuroimaging studies of the integration of prosodic information (e.g. Hesling et al [54]) have shown that integration of pitch information occurs in the right hemisphere, specifically the right inferior pre-frontal gyrus, and further work could investigate whether the effect found in the present study is right-distributed.

However, the effect is perhaps rather late to be a general effect of pitch determination. An alternative is that the effect could be specific to the detection of L+H* and H* (or some superset of each), particularly since it is so close in time to the intonation appropriateness effect. This could be further investigated by comparing the effects of different pitch accents and pitches over this time window to determine which accents or pitches elicit more positive or more negative voltages.

6 Conclusions

The goal of this study was to test the predictions made by Steedman's theory, that L+H* LH% marks theme and H* LL% marks rheme, by investigating the effect of anomalous pitch accents on sentence processing, and in particular, the effect on semantic processing. The results of the experiment showed that inappropriate intonation has a small but significant effect, approximately coincident with the N400 time window, in support of Steedman's theory. Further work is needed to determine whether the effect applies to only one or both pitch accents.

The study also identified effects of pitch accent and expected information structure, both of which could be further investigated to determine specific effects of the L+H* and H* pitch accents.

References

[1] Steedman M. , *Semantics and implicature in the meaning of English intonation*, unpublished, 2006
[2] Ferreira F., *Prosody and performance in language production*, unpublished, 2006
[3] Silverman K. et al, *ToBI: A standard for labelling English prosody*, Proceedings of the Second International Conference on Spoken Language Processing, 1992
[4] Pierrehumbert J., *The phonology and phonetics of English intonation*, 1980, Doctoral dissertation, MIT, Cambridge, Mass.
[5] Liberman M., Pierrehumbert J., *Intonational invariance under changes in pitch range and length*, in *Language Sound Structure*, Aronoff M., Oerhle R. Eds.,1984
[6] Maeda S., *A characterization of American English intonation*, 1976, Doctoral dissertation, MIT
[7] Vaissière J., *Language-independent prosodic features*, in *Prosody: Models and Measurements*, Cutler A., Ladd D.R., Brown G. Eds.,1983
[8] Beckman M.E., Elam G.E., *Guidelines for ToBI Labelling*, 1997
[9] Taylor P. , *Analysis and synthesis of intonation using the tilt model*, Journal of the Acoustical Society of America, 2000
[10] Calhoun S., *Phonetic dimensions of intonational categories - the case of L+H* and H**, Proceedings of Prosody, 2004
[11] Ladd R.D., Schepman A. , *'Sagging transitions' between high pitch accents in English: experimental evidence*, Journal of Phonetics, 2003
[12] Kruijff-Korbayova I., Steedman M. , *Discourse and information structure*, Journal of Logic, Language and Information, 2003
[13] Halliday M. , *Notes on transitivity and theme in English, Part II*, Journal of Linguistics, 1967
[14] Vallduví E., Vilkuna M., *On Rheme and Kontrast*, in *Syntax and Semantics, Vol. 29: The Limits of Syntax*, Culicover P., McNally L. Eds.,1998
[15] Gundel J.K., Fretheim T., *Topic and focus*, in *The Handbook of Contemporary Pragmatic Theory*, Horn L., Ward G. Eds.,2004
[16] Jackendoff R., *Semantic Interpretation in Generative Grammar*, 1972
[17] Pierrehumbert J., Hirschberg J., *The meaning of intonational contours in the interpretation of discourse*, in *Intentions in Communication*, Cohen P., Morgan J., Pollack M. Eds.,1990
[18] Steedman M. , *Information structure and the syntax-phonology interface*, Linguistic Inquiry, 2000
[19] Vallduví E., Engdahl E. , *The linguistic realization of information packaging*, Linguistics, 1996
[20] Gundel J.K., *On different kinds of focus*, in *Focus: Linguistic, Cognitive and Computational Perspectives*, Bosch P., van der Sandt R. Eds.,1999
[21] Hedbarg N., Sosa J.M., *The prosody of topic and focus in spontaneous English dialogue*, in *Topic and Focus: Papers from a Workshop on Intonation and Meaning*, Buring D., Gordon M., Lee C. Eds.,2002
[22] Watson D.G., Tanenhaus M.K., Gunlogson C.A., *Interpreting pitch accents in on-line comprehension: H* vs L+H**, MS University of Rochester, 2005
[23] Speckman E.-J., Elger C.E., *Introduction to the neurophysiological basis of the EEG and DC potentials*, in *Electroencephalography Basic Principles, Clinical Applications and Related Fields*, Niedermayer E., Lopes da Silva F. Eds.,1987
[24] Kutas M., Hillyard S.A. , *Reading Senseless Sentences: Brain Potentials Reflect Semantic Incongruity*, Science, 1980
[25] Kutas M., Hillyard S.A. , *Event-related potentials to semantically inappropriate and surprisingly large words*, Biological Psychology, 1980
[26] Kutas M., Hillyard S.A. , *Brain potentials during reading reflect word expectancy and semantic association*, Nature, 1984
[27] Connolly J.F., Phillips N.A., Stewart S.H., Brake W.G. , *Event-related potential sensitivity to acoustic and semantic properties of terminal words in sentences*, Brain and Language, 1992
[28] Besson M., Macar M. , *Visual and auditory event-related potentials elicited by linguistic and non-linguistic incongruities*, Neuroscience Letters, 1986
[29] Holcomb P.J., Coffey S.A., Neville H.J, *Visual and auditory sentence processing: a developmental analysis using event-related potentials*, in , Eds.,1992
[30] Kutas M., Hillyard S.A., *Event-related brain potentials to grammatical errors and semantic anomalies,* 1983

[31] Federmeier K.D., Kutas M., *Meaning and modality: influences of context, semantic memory organization, and perceptual predictability on picture processing*, 2001
[32] Osterhout L., Mobley L.A. , *Event-related brain potentials elicited by failure to agree*, Journal of Memory and Language, 1995
[33] Morris J., Holcomb P.J. , *Event-related potentials to violations of inflectional verb morphology in English*, Cognitive Brain Research, 2005
[34] Friederici A.D. , *Towards a neural basis of auditory sentence processing*, Trends in Cognitive Sciences, 2002
[35] Osterhout L., Holcomb P.J. , *Event-related potentials and syntactic anomaly*, Journal of Memory and Language, 1992
[36] Kaan E., Harris A., Gibson E., Holcomb P.J. , *The P600 as an index of syntactic integration difficulty*, Journal of Language and Cognitive Processes, 2000
[37] Osterhout L., Holcomb P.J., Swinney D.A., *Brain potentials elicited by graden-path sentences: evidence of the application of verb information during parsing*, in , Eds.,1994
[38] Osterhout L., Holcomb P.J., *Event-related potentials and syntactic anomaly: evidence of anomaly detection during the perception of continuous speech*, in , Eds.,1993
[39] Sutton S., Braren M., Zubin J., John E.R. , *Evoked potential correlates of stimulus uncertainty*, Science, 1965
[40] Donchin E., Coles M.G.H., *Is the P300 component a manifestation of context updating?*, 1988
[41] Näätänen R., *The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function*, Behavioral and Brain Sciences, 1990
[42] Steinhauer K., Alter K., Friederici A.D. , *Brain potentials indicate immediate use of prosodic cues in natural speech processing*, Nature Neuroscience, 1999
[43] Pannekamp A., Toepel U., Alter K., Hahne A., Friederici A.D. , *Prosody-driven sentence processing: An event-related brain potential study*, Journal of Cognitive Neuroscience, 2005
[44] Magne C., Astésano C., Lacheret-Dujour A., Morel M., Alter K., Besson M. , *On-line processing of "pop-out" words in spoken French dialogues*, Journal of Cognitive Neuroscience, 2005
[45] Heim S., Alter K., *Focus on focus: the brain's electrophysiological response to focus particles and accents in German*,
[46] Johnson S.M., Clifton C., Breen M., Morris Florak J., *An ERP investigation of prosodic and semantic focus*, Oralité e Gestulité, 2001
[47] Hruska C., Alter K., Steinhauer K., Steube A., *Misleading dialogues: Human brain's reaction to prosodic information*, Oralite e Gestulite, 2001
[48] Ito K., Garnsey S.M., *Brain responses to Focus-Related Prosodic Mismatch in Japanese*, Speech Prosody, 2004
[49] Toepel U., Alter K., *How mis-specified focus accents can distress our brain*, 15th ICPhS, 2003
[50] Heim S., Alter K. , *Prosodic pitch accents in language production: ERP data and acoustic analyses*, Acta Neurobiologiae Experimentalis, 2006
[51] Isel F., Alter K., Friederici A.D. , *Processing of syntactic and prosodic information in German particle verbs*, Journal of Cognitive Neuroscience, 2005
[52] American Electroencephalographic Society , *Guidelines for standard electrode position nomenclature*, Journal of Clinical Neurophysiology, 1994
[53] Semlitsch H.P., Anderer P., Schuster P., Presslich O. , *A solution for reliable and valid reduction of ocular artifacts, applied to the P300 ERP*, Psychophysiology, 1986
[54] Hesling I.H., Clément S., Bordessoules M., Allard M. , *Cerebral mechanisms of integration: Evidence from connected speech*, Neuroimage, 2005

7 Appendix I

This appendix contains sample materials for the experiment. The first four dialogues cover a single test case (the Anna/Manny case). The correct question/answer pairs are given here; the incorrect pairs are formed by swapping alternate answers (i.e. the incorrect question/answer pairs are question 1 with answer 2, question 2 with answer 1, question 3 with answer 4 and question 4 with answer 3.)

The same process is followed for different content; for brevity type 1 sentences only are currently listed below the full Anna/Manny case. Sentences are marked for pitch accent as spoken in the experimental materials.

1	Q: *I know EMMA will marry ARNIE. But who will marry MANNY?* *A: (ANNA) (will marry MANNY).* *H* L L+H* LH%*
2	*Q: I know EMMA will marry ARNIE. But what about ANNA?* *A: (ANNA) (will marry MANNY).* *L+H* LH% H* LL%*
3	*Q: I know EMMA will marry ARNIE. But who will ANNA marry?* *A: (ANNA will marry) (MANNY)* *L+H* LH% H* LL%*
4	*Q: I know EMMA will marry ARNIE. But what about MANNY?* *A: (ANNA will marry) (MANNY)* *H* L L+H* LH%*

Different content:	
1	*Q: I know AMY will run the MARATHON. But who will run the RELAY?* *A: (LEO) (will run the RELAY)* *H* L L+H* LH%*
2	*Q: I know ELLEN will buy the VAN. But who will buy the LORRY?* *A: (RYAN) (will buy the LORRY)* *H* L L+H* LH%*
3	*Q: I know LEIGH will learn MANDARIN. But who will learn MALAY?* *A: (RHONA) (will learn MALAY)* *H* L L+H* LH%*
4	*Q: I know MURRAY will dance the SAMBA. But who will dance the RHUMBA?* *A: (OWEN) (will dance the RHUMBA)* *H* L L+H* LH%*
5	*Q: I know LIAM kissed NORA. But who kissed LAURA?* *A: (OLLY) (kissed LAURA)* *H* L L+H* LH%*
6	*Q: I know LOUIS will play PIANO. But who will play VIOLIN?* *A: (MARY) (will play VIOLIN)* *H* L L+H* LH%*

7	Q: I know MARIO made the RAVIOLI. But who made the LAMB? A: (LYNN) (made the LAMB) H* L L+H* LH%
8	Q: I know BONNIE ate the ROLO. But who ate the LOLLY? A: (ALAN) (ate the LOLLY) H* L L+H* LH%
9	Q: I know LAUREL grew the MELON. But who grew the MARROW? A: (RORY) (grew the MARROW) H* L L+H* LH%
10	Q: I know the TIGER hunted the BUFFALO. But what hunted the RHINO? A: (The LION) (hunted the RHINO) H* L L+H* LH%
11	Q: I know PAUL will meet TAMMY. But who will meet MOLLY? A: (NORMAN) (will meet MOLLY). H* L L+H* LH%
12	Q: I know SAM will paint the WINDOW. But who will paint the WALL? A: (LORNA) (will paint the WALL) H* L L+H* LH%
13	Q: I know DANIELLE will bring ELEANOR. But who will bring LANA? A: (NEIL) (will bring LANA) H* L L+H* LH%
14	Q: I know GRAHAM will drink the COLA. But who will drink the LEMONADE? A: (MILLY) (will drink the LEMONADE) H* L L+H* LH%
15	Q: I know TOM will call RHIAN. But who will call LARRY? A: (NAOMI) (will call LARRY) H* L L+H* LH%
16	Q: I know HARRY drew the LEMON. But who drew the LIME? A: (ELLEN) (drew the LIME) H* L L+H* LH%
17	Q: I know CAMERON saw the BADGER. But who saw the MOLE? A: (LEON) (saw the MOLE) H* L L+H* LH%
18	Q: I know PAMELA will teach BEN. But who will teach MOIRA? A: (ROY) (will teach MOIRA) H* L L+H* LH%
19	Q: I know MATT will choose BLUE. But who will choose YELLOW? A: (EMILY) (will choose YELLOW) H* L L+H* LH%
20	Q: I know ELLA will visit IONA. But who will visit ISLAY? A: (MAUREEN) (will visit ISLAY) H* L L+H* LH%
21	Q: I know KELLY will thank JOHN. But who will thank IAN? A: (ELAINE) (will thank IAN) H* L L+H* LH%
22	Q: I know SHEILA dated LUKE. But who dated IRENE? A: (WARREN) (dated IRENE) H* L L+H* LH%
23	Q: I know JIMMY liked the BOOK. But who liked the FILM? A: (MARION) (liked the FILM) H* L L+H* LH%
24	Q: I know ERIC will carry the TABLE. But who will carry the WARDROBE? A: (RON) (will carry the WARDROBE) H* L L+H* LH%

25	*Q: I know KEN will help SARAH. But who will help RAY?* *A: (EILEEN) (will help RAY)* *H* L L+H* LH%*
26	*Q: I know WILLIAM planted the ROWAN. But who planted the LILY?* *A: (MONA) (planted the LILY)* *H* L L+H* LH%*
27	*Q: I know TOMMY will use the GARLIC. But who will use the ONION?* *A: (MIRIAM) (will use the ONION)* *H* L L+H* LH%*
28	*Q: I know WENDY will dig the COMPOST. But who will dig the MANURE?* *A: (NINA) (will dig the MANURE)* *H* L L+H* LH%*
29	*Q: I know LEANNE wrote the MUSIC. But who wrote the RHYME?* *A: (LYLE) (wrote the RHYME).* *H* L L+H* LH%*
30	*Q: I know COLIN will check the tyres. But who will check the OIL?* *A: (LAURIE) (will check the OIL)* *H* L L+H* LH%*
31	*Q: I know SIMON will clean the LOUNGE. But who will clean the HALL?* *A: (MURIEL) (will clean the HALL)* *H* L L+H* LH%*
32	*Q: I know ALICE will pay the JOINER. But who will pay the PLUMBER?* *A: (MARIE) (will pay the PLUMBER)* *H* L L+H* LH%*
33	*Q: I know JENNY will wear the BROOCH. But who will wear the RING?* *A: (LEILA) (will wear the RING)* *H* L L+H* LH%*
34	*Q: I know ANDY will walk the POODLE. But who will walk the COLLIE?* *A: (AARON) (will walk the COLLIE)* *H* L L+H* LH%*
35	*Q: I know ANGELA will survey MILAN. But who will survey ROME?* *A: (MELANIE) (will survey ROME)* *H* L L+H* LH%*
36	*Q: I know TANIA will pack the SHAMPOO. But who will pack the MIRROR?* *A: (WILMA) (will pack the MIRROR)* *H* L L+H* LH%*

8 Appendix 2

Total, mean and standard deviation (Stdev) of the number of accepted trials for each of the 11 subjects, by condition:

	First accented word							
	Early phrase boundary				Late phrase boundary			
	Rheme		Theme		Rheme		Theme	
Subject	H*	L+H*	H*	L+H*	H*	L+H*	H*	L+H*
1	34	34	37	37	37	36	35	33
2	35	36	35	37	31	35	35	35
3	37	34	33	36	36	37	36	37
4	31	30	30	30	33	26	31	30
5	36	37	35	35	35	37	35	36
7	37	37	37	34	35	35	36	36
9	35	33	36	36	33	35	35	36
10	31	27	27	31	31	28	27	29
11	37	36	36	36	36	37	36	36
12	36	35	35	37	35	35	37	34
14	32	36	35	35	35	35	37	34
Total	381	375	376	384	377	376	380	376
Mean	34.64	34.09	34.18	34.91	34.27	34.18	34.55	34.18
Stdev	2.34	3.11	3.09	2.39	2	3.68	2.98	2.6

Total, mean and standard deviation (Stdev) of the number of rejected trials for each of the 11 subjects, by rejection type:

Subject	Threshold	Drift	Artefact
1	4	24	0
2	6	26	0
3	0	19	0
4	0	107	0
5	2	25	0
7	5	12	0
9	8	30	0
10	0	115	0
11	11	5	0
12	12	13	3
14	21	17	2
Total	69	393	5
Mean	6.27	35.73	0.45
Stdev	6.47	37.94	1.04