The Nature of Theme and Rheme Accents

Sasha Calhoun

sasha@ling.ed.ac.uk

Centre for Speech Technology Research

October 9, 2003

Abstract

It has increasingly been recognised that appropriate intonation is essential to create believable voices for speech synthesis. This is particularly true in dialogue, where the link between intonation and meaning is especially important. Here we report two experiments, a production and perception study, which test an aspect of Steedman's (2000) theory relating information and intonation structure with a view to specifying intonation in a speech synthesis system. He claims that themes and rhemes, the basic building blocks of information structure, are marked by distinctive pitch accents in English, which he identifies with L+H* and H* in the ToBI system respectively. After reviewing problems with the identification of these ToBI accents, we show that speakers do produce and listeners do distinguish different pitch accents in these discourse contexts, but that the ToBI labels may not be helpful to characterise the distinction. The exact phonetic nature of theme and rheme accents remains unclear, but the alignment of the start of the rise, pitch height and the fall after the pitch peak all appear to be factors. Speakers also appear to be more sensitive to the distinction at the end of an utterance than utterance-medially.

1 Introduction

Natural-sounding and appropriate prosody is an important part of creating acceptable and understandable speech synthesis. This is particularly true in dialogue systems, where there is a clear link between prosody (which in this study will be taken to be the F0 contour) and meaning (e.g. Pierrehumbert & Hirschberg 1990). In order to make a believable machine dialogue agent, a way to link the semantics and the intonation of an utterance is needed.

Reliable techniques currently exist for synthesising natural-sounding intonation on the basis of a phonological representation such as the ToBI system (Clark 2003)(for a description of ToBI see Silverman, Beckman, Ostendorf, Wightman, Price, Pierrehumbert & Hirschberg 1992). Furthermore, there is a body of work mapping certain semantic concepts to particular ToBI intonational events (e.g. Pierrehumbert & Hirschberg 1990, Pierrehumbert & Steele 1989, Hedberg & Sosa 2001). Here, we adopt Steedman's (2000) theory, which claims that intonation structure is a partially specified representation of information structure. He claims that the primary division of a sentence into theme and rheme can be marked by distinctive pitch accents: L+H* and H*, on the head of the thematic and rhematic phrase respectively. However, the phonetic distinction between these two accents has proved problematic for ToBI annotators and it has been questioned whether they are, in fact, separate intonational categories (Taylor 2000). We report a pilot production study which seems to show that speakers do consistently produce distinctive theme and rheme pitch accents, and a perception study which shows that listeners are sensitive to the difference. The studies go some way to clarifying what the phonetic difference between the accents is, though this remains unsettled. We discuss the implications of these results for Steedman's theory and the usefulness of this and the ToBI annotation scheme in producing appropriate intonation contours in such dialogue contexts.

1.1 Information Structure and Intonation Structure

For the purposes of this study, we will assume a view of the relationship between information and intonation structure which draws on much previous work (e.g. Jackendoff 1972, Rooth 1992, Selkirk 1984, Buring to appear) but

has been most clearly formalised by Steedman (2000). During a dialogue, speakers build a model of the discourse made up of propositions about mutually agreed knowledge, in the manner of Grosz & Sidner (1986). For example, in this exchange:

(1) A: That guy's Henry Lombard, I think?
     B: That's Henry Lambert, not Henry Lombard.

A puts into the model the following proposition:

(2) $\exists x.is'x\ lombard'$

B, however, wants to replace this with this proposition:

(3) $\exists x.is'x\ lambert'$

The specification of information structure given in B's answer allows them to do this.

According to the theory, intonational events will be used to mark this structure. Intonation boundaries mark information structure boundaries, i.e. the partition of utterances into propositional units and then the division between thematic and rhematic information within each unit. Pitch accents mark the focus of each constituent and the type of pitch accent marks whether each element of a sentence is a theme or a rheme. Boundary tones are also said to have informational content, though this was not directly addressed in this study. Hence, B's reply above can be analysed in terms of information structure and corresponding intonational structure in the following way:

(4) ((That's) (Henry Lambert)), ((not) (Henry Lombard))
        theme   *rheme*   *rheme*   *theme*

(5) That's Henry Lambert, not Henry Lombard
       L+H*   H* LL%   H*   L+H* LH%

3

As Steedman is careful to point out, (5) represents a very careful pronunciation of the utterance. Often in real speech the intonational marking is underspecified. In particular, not all information structure boundaries are marked with intonation structure boundaries. Importantly, themes and rhemes are only likely to be marked with pitch accents if they are in focus. Steedman claims that this occurs when they are part of a salient alternative set in the discourse. For example, the theme in the first part of the sentence above is *that's*, or the man the speakers are referring to. This theme is in an alternate set of themes with anything else in the discourse domain the speakers could be referring to. However, since it is clear from A's question who the referent is, these alternatives are not salient and the pitch accent on *that* would be unlikely to be realised. In the second part of the sentence, however, the theme *Henry Lombard* stands in direct contrast to the alternative *Henry Lambert* and is therefore more likely to be accented. In this study we were interested in the phonetic description of these accents marking the theme and rheme, which we will call T and R accents respectively.

1.2   The L+H*/H* Controversy

As the previous section shows, in its current form Steedman's theory depends on there being a reliable phonetic difference between T and R accents, which he associates with the L+H* and H* accents in the ToBI system. According to the official ToBI annotation conventions, an H* accent is supposed to be given to peaks in the F0 contour associated with the stressed syllable (Beckman & Edwards 1992). An L+H* accent is also used to mark a peak, but one that is immediately preceded by a sharp rise from a valley in the lowest part of a speaker's pitch range. However, successive H* accents often occur with a F0 'dip' in between them. Pierrehumbert analyses this as a 'sagging transition' rather than an L, although she acknowledges this is not a very attractive solution to the problem. Many annotators have had difficulty distinguishing this 'dip' from an L target at the beginning of an L+H* accent. To make matters more confusing, annotators are instructed to treat the H* accent as a 'default' category, annotating any rising accent they are not sure of as H*. In a study of inter-annotator agreement in doing ToBI transcriptions, Pitrelli, Beckman & Hirschberg (1994) collapsed these categories because the level of disagreement would have swamped all their
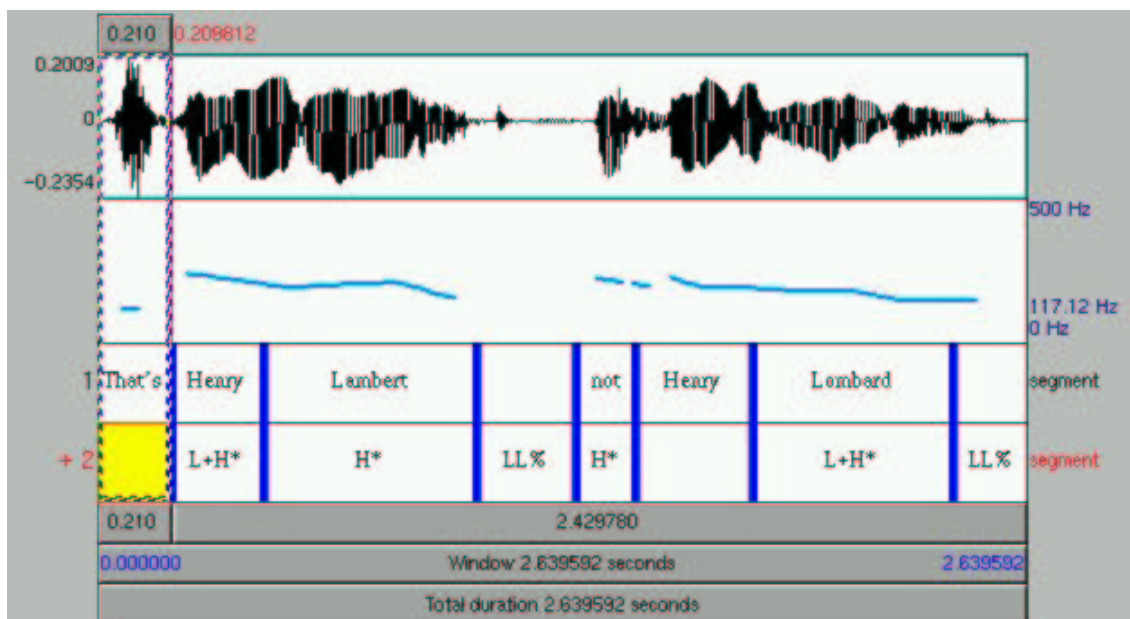
Figure 1: Example of the Difficulty in Distinguishing L+H* and H*

results in the other categories. In his Tilt intonation description system, which describes intonation events in terms of a continuous variable derived from the amount of rise and position in the accent, Taylor (2000) found that there is significant overlap in the coverage of the two accents. An example of the difficulty in analysing these accents can be seen in Figure 1.

Ladd & Schepman (2003) examined H* accents in a production study looking at pitch height and the alignment of L and H targets in relation to the segmental string in pairs such as *Norman Elson / Norma Nelson* and found that the 'sagging transition' before the accent was consistently aligned with the beginning of the accented syllable, implying an L target. They suggest that all accents with apparent L targets should be reanalysed as L+H*, reserving H* for peaks that start high. They allow, however, that there could be a difference between the accents (in terms of Pierrehumbert's phonological categories) in terms of the alignment of the L target in relation to the segmental string.

5

1.3   Present Research

We decided, as a result of these problems, that it would be difficult and inconclusive to test whether themes and rhemes are marked by L+H* and H* pitch accents respectively. Instead, we looked directly at whether there is a reliable phonetic difference between the pitch accents that mark themes and rhemes in contexts where these would be likely to occur, such as the example in (1) above.

The production study was exploratory, aiming to identify possible phonetic dimensions for the distinction between the accents, with a view to producing these accents in a speech synthesis system. We therefore concentrated on the height and alignment of Ls and Hs in relation to the segmental string. The perception study aimed to show more conclusively whether listeners are sensitive to the accent distinction and which of the phonetic dimensions were more important.

2   Experiment 1: Production Study

2.1   Method

In order to get as many reliable productions of T and R accents as possible in an appropriate phonetic environment we used read speech as it is much easier to control the both the informational context and the words used. Eight sentences were constructed. The target word was the head of the theme or rheme phrase and was at least two syllables long with the primary stress not on the last syllable, in order to separate the pitch accent from nearby boundary tones. The accented syllable in each target word began and ended with a sonorant consonant, providing a continuous F0 signal and making the speech signal easier to segment using a spectrogram.

Each sentence was presented in four versions, so that each target word would appear as both a theme and a rheme in both clauses of each sentence:

(6)   Q: That guy's Henry Lombard, I think?

A: That's Henry *Lambert*, not Henry *Lombard*.

(7)   Q: That guy's Henry Lombard, I think?
      A: That isn't Henry *Lombard*, it's Henry *Lambert*.

(8)   Q: That guy's Henry Lambert, I think?
      A: That's Henry *Lombard*, not Henry *Lambert*.

(9)   Q: That guy's Henry Lambert, I think?
      A: That isn't Henry *Lambert*, it's Henry *Lombard*.

These sentences were divided randomly into four blocks, so that there was not more than one version of each sentence in any block. They were presented to the speaker along with 24 distractor sentences, making four blocks of 14 sentences each, or 56 sentences in total (see Appendix A for full list). This made a potential 32 tokens of each of the T and R accents.

One speaker, an undergraduate at the University of Edinburgh, was used for her ability to consistently produce well-modulated, natural-sounding speech when reading aloud. In a sound-proofed recording studio, the author asked the speaker each question in turn and our speaker replied. The dialogues were recorded digitally.

The target words (e.g. *Lombard* and *Lambert* above) were then analysed using *xwaves* (Entropic-Research-Labs 1998). The author determined, by listening to the recording and looking at the pitch track, whether the accented syllable in each target word was associated with a clear pitch movement. If it was, then, using the audio, pitch track, wave form and spectrogram associated with each word, the author labelled the following points in each accent:

1. C0: the beginning of the consonant of the stressed syllable

2. V0: the beginning of the vowel of the stressed syllable

3. C1: the beginning of the consonant following the stressed vowel

4. V1: the beginning of the vowel following the stressed vowel

5. L: the pitch low point, or point where the pitch track begins to rise sharply, before the pitch accent

		C0	L	V0	H	C1	V1	T0-T1
F0 (Hz)	T accent	166.8	183.5	177.7	227.5	217.6	166.4	8.1
	R accent	210.0	208.1	232.4	268.7	260.4	186.9	54.2
Time (secs)	T accent	-0.059	-0.001	0.000	0.097	0.084	0.209	-
	R accent	-0.059	-0.053	0.000	0.101	0.083	0.199	-

Table 1: Results from Experiment 1

6. H: the pitch peak, or the turning point of the pitch track at the height of the pitch accent

7. T0: the f0 level at the intensity peak in the last syllable in the word before the target one

8. T1: the f0 level at the intensity peak in the syllable following the accented one (but before any boundary tone rise, if present)

2.2 Results

Of the 32 theme tokens, 7 were judged by the author to have been produced with a clear pitch accent. 29 of the 32 rheme tokens were produced with a clear pitch accent. This result in itself indicates it is not just the salience of an alternative set which means that a theme is marked with a pitch accent. But as, in this study, we were concerned with the realisation of pitch accents, the unaccented productions were put aside. Each of the seven T pitch accent tokens was matched with its corresponding R pitch accent token, and the remainder of the R tokens were excluded from analysis.

Table 1 shows the results from Experiment 1, where the labels are as described in the hypothesis above. Times are normalised relative to V0, which is taken to be 0 seconds.

As can be seen, these results seem to support the segmental alignment difference between the two accents suggested by Ladd & Schepman (2003) above. For the T accent, L is aligned with V0; whereas the R accent begins to rise

earlier, at C0. This result is highly significant using a two-tailed paired t-test ($P < 0.004$). H, however, seems to be aligned a short way into the next consonant for both accents.

The results also suggest that the difference between the two accents could be indicated by pitch height. Both L and H were produced with lower F0 for T accents than for R accents. These results only tended towards significance ($p < 0.129$ and $p < 0.112$ using a two-tailed paired t-test respectively), however the sample size was small. R accents also seemed to be followed by a dip in F0, to well below the starting F0 level; whereas the F0 level after a T accent seems to return to approximately its starting point. This result was significant (for $T0 - T1$, $p < 0.016$ using a two-tailed paired t-test).

The results of our production study show that in utterances where our speaker produced pitch accents in the expected places (i.e. on the head of the theme or rheme phrase), she consistently produced T and R pitch accents differently. It was not clear, however, whether it was the alignment differences, the relative pitch levels or a combination of these forming the overall shape of the two accents which signals the two accents to a hearer.

3   Experiment 2: Perception Study

The second experiment, a perception study, tested which of the hypothesised differences between T and R accents, if any, are perceptible. Listeners were presented with a forced-choice exercise. Subjects heard two versions of the dialogues outlined above, with the pitch accent on the theme having been altered, and were asked to choose which dialogue they thought was more natural-sounding. There were two main hypotheses:

1. Subjects would prefer dialogues in which the pitch accent on the theme was produced with a T accent to dialogues where the theme was produced with an R accent.

2. Subjects would prefer dialogues in which the the pitch accent on the theme was produced when each of four parameters (Alignment, Height,

Fall and Boundary[1]) was in the 't' setting, rather than the 'r' setting.

3.1 Method

The recordings from the first experiment were used to generate the stimulus materials. Four sentence types were used (1.7, 1.11, 1.12 and 1.14 in Appendix A and their variants). The questions were played back as they were recorded. The pitch tracks of the answers were manipulated and resynthesised using *Praat* tools (Boersma & Weenink 2003). Firstly, the pitch track of the entire answer was stylised automatically so that it was represented visually by straight lines drawn between pitch points at key turning points (approximately 15 points per utterance) using a *Praat* tool. Then the location of these pitch points was altered manually so that there was a pitch point at relevant points (C0, V0, H, T1, B0 and B1, see below) in the pitch accent on the theme.

A *Praat* script was then used to generate 16 versions of each sentence. Each version had its key pitch points altered to match with all possible combinations of each of the following four parameters in each of their two hypothesised settings ('t'-like and 'r'-like), see Figure 2. The sentence was then resynthesised with the altered pitch track using the PSOLA technique with a different *Praat* tool. Values used were decided on the basis of the production study. Ratios were used rather than absolute differences in F0 values as this is closer to human perception of pitch (see Ladd 1996, chap.7).

1. Alignment: 't': set time of L at V0, 'r': set time of L at C0;

2. Height: Set time of H 20% into following C. Set F0 of L to be 20% less than H. 't': set H to be 210Hz, 'r': set H to be 250Hz;

3. Fall: Set time of T1 at a stable point in the vowel following the accented one. 't': set F0 of T1 to be 10% lower than L, 'r': set F0 of T1 to be 20% lower than L;

---

[1]This factor was included as, on inspection, a disproportionately high number of themes were found to be followed by rising boundaries. It did not prove to be significant in the perception study, however, and so is not discussed.
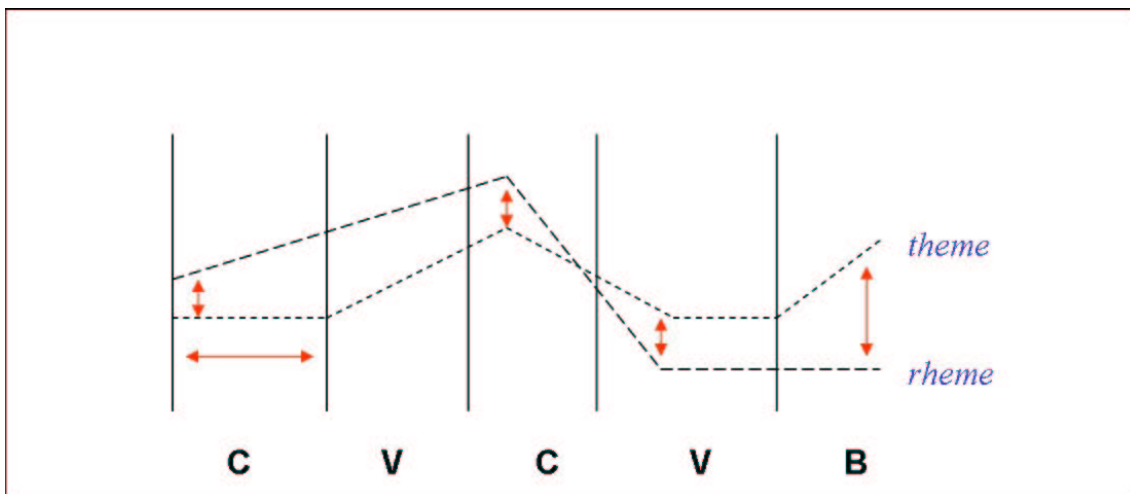
Figure 2: Pitch Accent Shape Variations Used in Perception Experiment

4. Boundary: Set time of B0 50ms before end of phrase, F0 same as T1. Set time of B1 at end of phrase. 't': set B1 to be 20% higher than B0, 'r': set B1 to be the same as B0.

These answers were then used to set up pairs of dialogues for subjects to choose between. For the first hypotheses we paired answers that differed by three parameter settings (i.e. either 4 't'-like versus 1 't'-like (4-1) or 3 't'-like versus 0 't'-like (3-0), assuming that these were equivalent). The second hypothesis was tested by pairing answers that differed only by each one of the four parameters in turn (i.e. either 3-2 or 2-1).

Both versions ('It isn't A, it's B' (theme-rheme) and 'It's B, not A' (rheme-theme)) of each of the four answers were tested with each of the 16 resulting parameter pairings. In addition, each pairing was tested in both orders, so subjects heard both the 'good' dialogue before the 'bad' dialogue and vice versa. This was because it has been found in previous forced-choice intonation studies that subjects have a preference for the version they hear most recently (e.g. Chorianopoulou 2002). This made a total of 256 dialogue pairs. These were divided evenly into 16 blocks of the 16 dialogue pairs

each so that each of the 16 parameter settings appeared once in each block and there were two instances of each version of each dialogue in each block. Within each block dialogues appeared in random order.

In order to keep the experiment to a reasonable length, half the blocks were presented to half the subjects and half to the other half. Therefore, in each of the five experiments (one for hypothesis 1, four for hypothesis 2), subjects was the random factor. There were four within-subjects factors: Sentence (Lombard, London, Malaya, Wombats); Place (theme-rheme or rheme-theme sentence); Order (good-bad or bad-good); and Type (the parameter pairing used - four combinations in hypothesis 1 and three in hypothesis 2).

Thirty subjects, staff and students at the University of Edinburgh, took part in the experiment in return for a small monetary reward. Subjects were told that they would hear two dialogues, and that the intonation contour of the answer would be different in each one. They were asked to choose which answer sounded like a more natural response to the question. Subjects began with a practice block consisting of 16 4-1 and 3-0 sentences not in the main experiment. They then heard eight blocks of dialogues, with a break after every two blocks. Subjects were told there was no time pressure in responding but that the dialogues could not be repeated. The entire session took about 45 minutes.

3.2   Results

In relation to the first hypothesis, it was found that subjects did prefer answers produced with a T accent on the theme to answers with an R accent on the theme. Overall 66.7% chose the 4-1 and 3-0 sentences with more 't' settings. This was significantly more than chance (using a 2 x 2 chi squared test, $\chi^2 = 115.8$, $p < 0.01$). However, this result was affected both by the order in which the stimuli were presented ('good'-'bad' and 'bad'-'good') and by the type of sentence (theme-rheme and rheme-theme). Using a 1 x 2 repeated-measures ANOVA there was a significant main effect of Order, $F(1, 24) = 6.508$, $p = 0.018$; similarly for Place, $F(1, 24) = 4.617$, $p = 0.042$. These two variables seemed to interact, though not significantly, $F(1, 24) = 2.062$, $p = 0.164$. This can be seen in Figure 3. For the theme-rheme ordered answers, the order of presentation was significant. When subjects
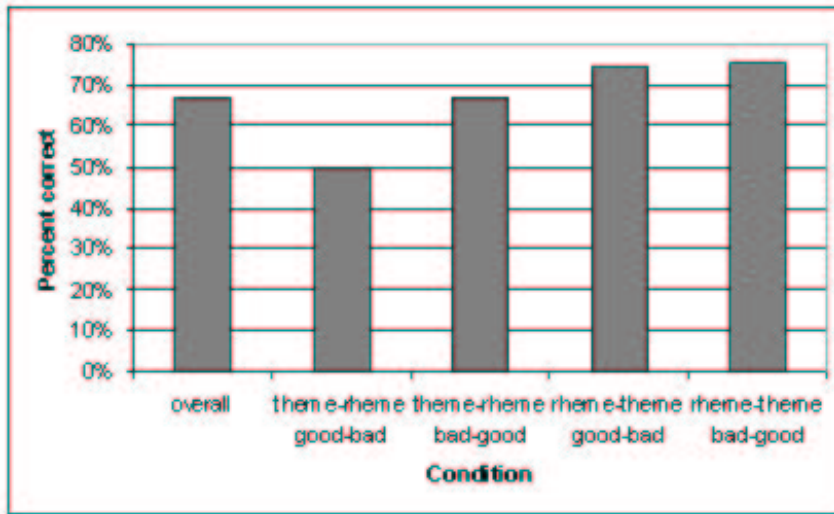
Figure 3: Effect of Different Test Conditions on Hypothesis 1 Results

heard the 'good' (more theme-like) version second, they preferred it 66.9% of the time, whereas when the good version was presented first, they performed only at the level of chance. For the rheme-theme ordered sentences, however, subjects reliably preferred the more theme-like version in either order.

Subjects performed better on the *Lombard-Lambert* sentences (1.14 and variants in Appendix A), and worse on the *monkeys-wombats* sentences (1.11 and variants in Appendix A), 82.3% and 56.5% respectively, although there was no significant main effect of Sentence, $F(3, 22) = 1.195$, $p = 0.335$. Many subjects commented on the strangeness of the *monkeys-wombats* sentences, suggesting that the preference for T accents on themes in appropriate and familiar contexts is even stronger than these results suggest. Our assumption that inverse parameter settings could be treated as equal (e.g. that 4-1 is effectively the same as 3-0) proved to be justified, 67.4% and 65.9% respectively, there is no main effect of type.

In relation to the second hypothesis, the only single parameter which caused subjects to significantly prefer the answer with that parameter in its 't' set-

ting on the theme pitch accent was Height. 73.4% of subjects chose the version of the answer in this category with lower pitch. This was significantly better than chance (using a 2 x 2 chi squared test $\chi^2 = 140.8$, $p < 0.01$). Within this category, subjects performed slightly better if both the Alignment and the Fall were also in their 't' setting, although this was not significant. This may show that though pitch height is the most robust indicator of the T accent, alignment and the fall may also be acoustic cues, though they are not sufficiently strong to indicate the accent on their own.

Overall, then, this experiment showed that the phonetic differences between T and R accents shown in the production study are perceptible. People, in production and perception, expect to hear a T accent in the theme position, at least in the sentence context used in these experiments. On the basis of the second experiment, it seems that the nature of the pitch accent on the theme at the end of a sentence is more important than at the beginning. The second experiment was not conclusive as to the exact phonetic distinction between T and R accents. Pitch height appears to be an important factor, however there are indications that other factors, particularly the fall after the pitch accent, may be important even though they are not robust enough to signal the pitch accent on their own.

4   Discussion and Conclusion

These results seem to show Steedman's claim that themes and rhemes are marked by distinctive pitch accents is correct. People are sensitive to the type as well as the presence of particular pitch accents in certain discourse contexts. It is not simply enough, as previous researchers have suggested (see review in Cutler, Dahan & van Donselaar 1997), to say that all themes are unaccented and all rhemes accented (usually equated with given and new). Nor does it seem to be the case that 'contrastive themes' are a special case of focus and will be acceptable as long as they are marked by a rising pitch accent, as seems to be suggested by Rooth (1992). In terms of using this finding to predict the placement of T accents in a speech synthesis system, however, we need to have more of an idea what factors influence whether a theme is realised with a pitch accent. Our pilot study shows that it is not simply the availability of alternative sets.

14

The implication in Steedman's work, and indeed the ToBI system, is that the interpretation of these pitch accents is entirely compositional. This could be brought into question by our results (c.f. Grabe, Gussenhoven, Haan, Marsi & Post 1998). People are more sensitive, and in fact if the order of presentation effect is taken into account, may only be sensitive, to the type of pitch accent on the theme at the end of the sentence. Further investigation is needed to see whether this is a semantic or a phonetic effect. The rheme-theme ordering could be more marked and therefore has a different interpretation than the theme-rheme ordering. This is plausible as the rheme-theme ordering within a clause is relatively unusual cross-linguistically (e.g. Vallduvi & Vilkuna 1998). Or it could simply be that listener's memory for a pitch accent in the middle of an utterance is less reliable than that at the end of an utterance.

A third possibility is that the phonetic description of an L+H* H* sequence is not the reverse of a H* L+H* sequence, which would seem to follow logically from compositionality. Perhaps the most worrying finding in terms of using ToBI to describe intonation for speech synthesis is that the most reliable cue to the two accents was pitch height, a factor which is not built into the description of the two accents at all[2]. Pitch height is a difficult factor to model independently of the overall F0 contour, as it interacts with general F0 declination during an utterance (which would make non-compositionality likely), paralinguistic factors such as uncertainty and emphasis (Pierrehumbert & Steele 1989, Ladd & Morton 1997) and extralinguistic factors such as the identity of the speaker. However, if, as this study seems to show, it is encodes semantic distinctions, the intonation annotation needs to capture this and the synthesis system needs to be able to model it.

These findings give further evidence that the L+H*/H* annotation does not seem to be propitious in distinguishing the phonetics of the T and R accents. Taking both phonetic and phonological considerations into account, there appear to be three classes of accents in this range: two starting with L targets that can be distinguished on phonological grounds (which we are calling T and R here), and one with no L target that can be distinguished on phonetic grounds (which always seems to be an R accent). In trying to distinguish the former two phonetically, the alignment difference suggested by Ladd & Schepman (2003) seems to be too fine to act as a reliable cue for

---

[2]Although, of course, alignment of the L tone is not included either, though this stipulation seems a rather more compatible addition to the present scheme

listeners. In addition, a more robust cue to these accents, the depth of the fall after the accent, is not encoded in the ToBI system at all. Lastly, our pilot work suggested that the difference may lie in the curve of the rise before the accent, with R accents rising more steeply than T accents. This again is not part of the ToBI system and it is not clear how this can be reliably captured in a phonetic description. To suggest three accents in this space, however, seems implausible on perceptual grounds and it is unclear how to resolve the conflicting phonological and phonetic considerations.

Our production and perception studies show that speech synthesis systems, particularly in dialogue, do need to pay attention to the nature as well as the location of pitch accents. We have shown that people both produce and perceive a difference between pitch accents on themes and rhemes. The exact nature of these accents remains unclear, but we have further questioned their identification as L+H* and H*, in part because pitch height appears to be important in their description.

References

Beckman, M. & Edwards, J. (1992), Intonational categories and the articulatory control of duration, *in* E.-B. Tohkura & Y. Sagisaka, eds, 'Speech Perception, Production and Linguistic Structure', OHM Publishing Co. Ltd., pp. 356–375.

Boersma, P. & Weenink, D. (2003), 'Praat:doing phonetics by computer', http://www.praat.org.

Buring, D. (to appear), 'On d-trees, beans and b-accents', *Linguistics and Philosophy* .

Chorianopoulou, E. (2002), Evaluating prodosy prediction in synthesis with respect to modern greek prenuclear accents, Master's thesis, Linguistics, University of Edinburgh.

Clark, R. (2003), Generating Synthetic Pitch Contours Using Prosodic Structure, PhD thesis, University of Edinburgh, Edinburgh, UK.

Cutler, A., Dahan, D. & van Donselaar, W. (1997), 'Prosody in the comprehension of spoken language: A literature review', *Language and Speech* **40**(2), 141–201.

Entropic-Research-Labs (1998), '*xwaves* manual'. version 5.3.1.

Grabe, E., Gussenhoven, C., Haan, J., Marsi, E. & Post, B. (1998), 'Preaccentual pitch and speaker attitude in dutch', *Language and Speech* **41**(1).

Grosz, B. & Sidner, C. (1986), 'Attention, intentions, and the structure of discourse', *Computational Linguistics* **12**, 175–204.

Hedberg, N. & Sosa, J. M. (2001), The prosodic structure of topic and focus in spontaneous english dialogue, *in* 'LSA Summer Institute Workshop on Topic and Focus', Santa Barbara.

Jackendoff, R. S. (1972), *Semantic Interpretation in Generative Grammar*, MIT Press, Cambridge, MA.

Ladd, D. R. (1996), *Intonational Phonology*, Cambridge University Press, UK.

Ladd, D. R. & Morton, R. (1997), 'The perception of intonational emphasis: continuous or categorical?', *Journal of Phonetics* **25**, 313–342.

Ladd, D. R. & Schepman, A. (2003), "sagging transitions' between high pitch accents in English: experimental evidence', *Journal of Phonetics* **31**(1), 81–112.

Pierrehumbert, J. & Hirschberg, J. (1990), The meaning of intonational contours in the interpretation of discourse, *in* P. Cohen, J. Morgan & M. Pollack, eds, 'Intentions in Communication', MIT Press, Cambridge, MA, pp. 271–311.

Pierrehumbert, J. & Steele, S. (1989), 'Categories of tonal alignment in english', *Phonetica* **46**, 181–196.

Pitrelli, J., Beckman, M. & Hirschberg, J. (1994), Evaluation of prosodic transcription labelling reliability in the ToBI framework, *in* 'Proceedings of the Third International Conference on Spoken Language Processing', Vol. 2, pp. 123–126.

Rooth, M. (1992), 'A theory of focus intepretation', *Natural Language Semantics* **1**, 75–116.

Selkirk, E. (1984), *Phonology and Syntax*, MIT Press, Cambridge, MA.

Silverman, K., Beckman, M., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J. & Hirschberg, J. (1992), A standard for labelling english prosody, *in* 'Proceedings of the International Conference on Spoken Language Processing (ICSLP)', Vol. 2, Banff, pp. 867–870.

Steedman, M. (2000), 'Information structure and the syntax-phonology interface', *Linguistic Inquiry* **31**(4), 649–689.

Taylor, P. (2000), 'Analysis and synthesis of intonation using the tilt model', *Journal of the Acoustical Society of America* **107**, 1697–1714.

Vallduvi, E. & Vilkuna, M. (1998), 'On rheme and kontrast', *Syntax and Semantics* **29**, 79–108.

A Appendix A: Sentences for Production Experiment

A.1 Section

1. Q: Don't you have to be very fit to climb Ben Nevis?
   A: No, Ben Nevis is an easy climb.

2. Q: Isn't that book by Alan Lowry?
   A: It's by Anna Lowry, not by Alan Lowry.

3. Q: What method did the psychiatrist use?
   A: He has tried a course of hypnosis.

4. Q: That's Jane Vanderberg, isn't it?
   A: It's not Jane Vanderberg, it's Jane Mulder.

5. Q: That's money laundering you're suggesting!
   A: It's just a financial solution to the problem, not money laundering.

6. Q: Do you think 'The Matrix' was an arthouse or an indie film?
   A: I don't know, I haven't seen 'The Matrix'.

7. Q: Where is her place again? In Longmore?
   A: It isn't in Longmore, it's in London.

8. Q: Which is the coldest month of the year?
   A: Probably either January or February.

9. Q: That piece comes from Norma Munroe, doesn't it?
   A: It's not from Norma Munroe, it's from Norman Munroe.

10. Q: Who was in charge of planning the scheme?
    A: Jeremy McConville headed the team.

11. Q: Didn't you tell me that she had some monkeys?
    A: I didn't know she had some monkeys, I knew she had some wombats.

12. Q: She's from Havana, isn't she?
    A: She's from Malaya, not from Havana.

13. Q: What are the common symptoms of chicken pox?
    A: Red dots on the skin are common signs.

14. Q: That guy's Henry Lambert, I think.
    A: That's Henry Lombard, not Henry Lambert.

A.2   Section

1. Q: I'm just suggesting a financial solution to the problem...
   A: It isn't a financial solution to the problem, it's money laundering.

2. Q: Henri plays for Arsenal not Leeds, doesn't he?
   A: Yeah, he plays for Arsenal.

3. Q: Where is her place again? In London?
   A: It's in Longmore, not in London.

4. Q: Where does organic food come from?
   A: It comes from Greenock.

5. Q: She's from Malaya, isn't she?
   A: She isn't from Malaya, she's from Havana.

6. Q: Have you seen Jim lately?
   A: No, Jim's doesn't live in Edinburgh anymore.

7. Q: That guy's Henry Lombard, I think.
   A: That isn't Henry Lombard, it's Henry Lambert.

8. Q: What do you think is the best mountain to climb in Scotland?
   A: Ben Nevis is one of the best and most managable.

9. Q: Isn't that book by Anna Lowry?
   A: It's not by Anna Lowry, it's by Alan Lowry.

10. Q: Who had on their new high heels?
    A: Kate was wearing her new Jimmy Choos.

11. Q: That's Jane Mulder, isn't it?
    A: It's Jane Vanderberg, not Jane Mulder.

12. Q: Do you think the weather's worst in January?
    A: No, I think February can be more bitter.

13. Q: That piece comes from Norman Munroe, doesn't it?
    A: It comes from Norma Munroe, not from Norman Munroe.

14. Q: Didn't you tell me that she had some wombats?
    A: I thought she had some monkeys, not some wombats.

A.3   Section

1. Q: That guy's Henry Lambert, I think.
   A: That isn't Henry Lambert, it's Henry Lombard.

2. Q: Where is Jim from originally?
   A: Jim's from Edinburgh.

3. Q: She's from Havana, isn't she?
   A: She isn't from Havana, she's from Malaya.

4. Q: Why do adults have to be wary of red dots on the skin?
   A: Chicken pox as an adult can be deadly.

5. Q: Where is her place again? In Longmore?
   A: It's in London, not in Longmore.

6. Q: Do you think it's worth trying hypnosis?
   A: I don't think hypnosis is worthwhile.

7. Q: That's money laundering you're suggesting!
   A: It's not money laundering, it's just a financial solution to the problem.

8. Q: Who does Henri play for?
   A: He plays for Arsenal.

9. Q: Didn't you tell me that she had some monkeys?
   A: I knew she had some wombats, I didn't know she had some monkeys.

10. Q: That piece comes from Norma Munroe, doesn't it?
    A: It comes from Norman Munroe, not from Norma Munroe.

11. Q: What days are the classes run?
    A: The classes are run on Mondays, Wednesdays and Fridays.

12. Q: That's Jane Vanderberg, isn't it?
    A: It's Jane Mulder, not Jane Vanderberg.

13. Q: What was Jeremy's role in the process?
    A: Jeremy McConville headed the management team.

14. Q: Isn't that book by Alan Lowry?
    A: It's not by Alan Lowry, it's by Anna Lowry.

A.4   Section

1. Q: Didn't you tell me that she had some wombats?
   A: I didn't know she had some wombats, I thought she had some monkeys.

2. Q: What time is the movie, 8 o'clock? A: No, the movie starts at 9 o'clock tonight.

3. Q: That piece comes from Norman Munroe, doesn't it?
   A: It isn't from Norman Munroe, it comes from Norma Munroe.

4. Q: Why did Kate look so sad last night?
   A: She broke her new Jimmy Choos.

5. Q: That's Jane Mulder, isn't it?
   A: It's not Jane Mulder, it's Jane Vanderberg.

6. Q: Why is Greenock popular with hippies?
   A: Organic food comes from Greenock.

7. Q: Isn't that book by Anna Lowry?
   A: It's by Alan Lowry, not by Anna Lowry.

8. Q: What's your favourite film of the past few years?
   A: Definitely 'The Matrix'.

9. Q: That guy's Henry Lombard, I think.
   A: That's Henry Lambert, not Henry Lombard.

10. Q: She's from Malaya, isn't she?
    A: She's from Havana, not from Malaya.

11. Q: What days does Barry have off work?
    A: Barry doesn't work on Mondays, Wednesdays and Fridays.

12. Q: Where is her place again? In London?
    A: It isn't in London, it's in Longmore.

13. Q: Do you agree with his suggestion to use hypnosis?
    A: No, I think meditation is a better treatment than hypnosis.

14. Q: I'm just suggesting a financial solution to the problem...
    A: It's money laundering, not a financial solution to the problem.