Phonetic Dimensions of Intonational Categories - the case of L+H* and H*

Sasha Calhoun

Centre for Speech Technology Research,
University of Edinburgh, UK

Sasha.Calhoun@ed.ac.uk

Abstract

ToBI, in its conception, was an attempt to describe intonation in terms of phonological categories. An effect of the success of ToBI in doing this has been to make it standard to try to characterise all intonational phonological distinctions in terms of ToBI distinctions, i.e. segmental alignment of pitch targets and pitch height as either High or Low. Here we report a series of experiments which attempted to do this, linking two supposed phonological categories, theme and rheme accents, to two controversial ToBI pitch accents L+H* and H* respectively. Our results suggest a reanalysis of the dimensions of phonological intonational distinctions. It is suggested that there are three layers affecting the intonational contour: global extrinsic, local extrinsic and intrinsic; and the theme-rheme distinction may lie in the local extrinsic layer. It is the similarity both of the phonetic effects and the semantic information conveyed by the last two layers that has led to the confusion in results such as those reported here.

1. Introduction

Since the ToBI transcription system became standard to describe intonation [8], it has also been common practice to think of the only categorical intonational distinctions as being those described in the ToBI system, in particular the alignment of pitch accents with stressed syllables and pitch height as either High or Low [7].[1] These categorical distinctions are meant to be phonologically justifiable by their link to certain semantic notions [6, 9]. However, a number of recent studies have shown that the phonetic description of phonological intonation categories may involve more segmental alignment information than just association of peaks and troughs with stressed syllables (see [5]) and that there may be more categorical linguistic divisions than High and Low within the pitch scale [4]. Broadly, the phonetic dimensions of intonational categories are open to debate. It may be that the phonetic description of ToBI just needs to be fine-tuned, or it may be that ToBI is not adequate to describe certain real intonational phonological categories.

In a recent series of production and perception experiments, we looked at one claim that links a semantic distinction - the division of a sentence into theme and rheme - to two ToBI pitch accents, L+H* and H* respectively [9]. The phonetic specification of these accents has caused a lot of controversy, being argued to either not exist at all or be wrongly drawn in the ToBI specifications (discussed in [5]). This is because many H* accents have an apparent L target at the start of their rise and because the distinction is also sometimes informally held to involve peak height (with the H* in L+H* being lower). In a controlled production study involving read sentence contexts in which we thought it likely that the theme and rheme would be marked with a pitch accent, we found that the two accents both had apparent L targets, and were distinguished in terms of alignment of this L to the segmental string, by the height of the peak and by the strength of the fall after the peak. However, in a complementary perception study, the only single factor strong enough to signal this difference to listeners was peak height.

In this paper, I offer a different analysis of these results, arguing that we must look again at the import of relative pitch levels in the semantic interpretation of intonation. Drawing on a proposal by Ladd [3, chap.7], which is very similar to Bolinger [1], I propose there are three separate domains which influence the intonation (by this we mean here F0) of an utterance. Ladd calls these global extrinsic, local extrinsic and intrinsic effects. I propose that the last two have a clear effect on the semantics of utterances, with the theme-rheme distinction operating at the local extrinsic level. The previous difficulties in characterising the distinction were in fact caused by an interaction between the acoustic signals to the last two layers. There is a definite need for more concrete research into both the semantics and phonetics of this intermediate layer.

[1] We leave aside the identification and semantics of boundary tones for the purposes of this paper.

2. Theme and Rheme Accents: Two Experiments

2.1. Production Study

Our two experiments aimed to test whether there is a reliable phonetic difference between the pitch accents that mark themes and rhemes in discourse contexts where they would be likely to occur, such as (1) below.[2] We began by looking at the kinds of phonetic correlates that are said to mark the distinction between L+H* and H* (although these phonetic correlates are not crucial to Steedman's claim).

(1)	(That's Henry Lambert), *rheme*	(not Henry Lombard) *theme*
(2)	That's Henry Lambert, H* LL%	not Henry Lombard L+H* LH%

2.1.1. *Method*

Eight similar sentences were constructed. In each case the target word was phonetically suitable to get a continuous F0 signal and a pitch movement that would be separate from nearby boundary tones. Each sentence was presented in four versions, so that each target word would appear as both a theme and a rheme

[2] It is also possible that this sentence could be analysed as having *that* as a theme and *not* as a rheme. However, this was not crucial here as we were only manipulating the accents on *Lambert* and *Lombard*.

Table 1: *Results from Production Experiment*

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

in both clauses of each sentence. The sentences were ordered randomly and presented to the speaker along with 24 distractor sentences in four blocks of 14 sentences each, 56 sentences in total. 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 produce natural-sounding speech when reading aloud. In a sound-proofed recording studio, the author asked the speaker each question in turn and the speaker replied.

It was then determined, by listening to the recording and looking at the pitch track, whether 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, key points were labelled in each accent as indicated in Figure 1.

2.1.2. *Results*

Of the 32 theme tokens, 7 were judged 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 that there is more to the marking of themes than a certain pitch movement, but as this was not the central concern of the study, 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 this experiment, where the labels are as described in the hypothesis above. Times are normalised relative to V0, which is taken to be 0 seconds.

These results seem to support the segmental alignment difference between the two accents suggested by Ladd & Schepman [5]. For the T accent, L is aligned with V0; whereas the R accent rises earlier, at C0. This result is highly significant using a two-tailed paired t-test ($P < 0.004$). The results also suggest there could be a pitch height difference. 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 significantly greater dip in F0 than T accents (for $T0 - T1, p < 0.016$ using a two-tailed paired t-test).

2.2. Perception Study

The perception study tried to test firstly whether listeners could perceive the difference between T and R accents (as determined by the production experiment) (Hypothesis 1); and secondly which, if any, of the four factors (Alignment, Height, Fall and Boundary[3]) would they be sensitive to (Hypothesis 2). Listeners

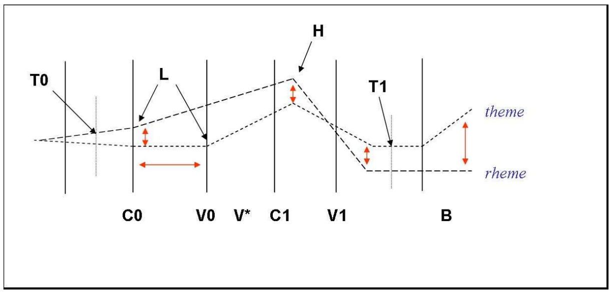

Figure 1: *Pitch Accent Labelling used in Production Study and Shape Variations Used in Perception Study (vertical lines mark segment boundaries, C0 is beg. of first consonant in stressed syllable, V0 is beg. of first vowel in stressed syllable, C1 and V1 are the equivalent in the next syllable, V* is the stressed vowel, T is at the vowel mid-point, B=boundary)*

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 resynthesised, and were asked to choose which dialogue they thought was the more natural and appropriate answer to the question.

2.2.1. *Method*

The recordings from the first experiment were used to generate the stimulus materials. Four sentence types were used. Each sentence was then resynthesised using the PSOLA technique to produce sixteen versions of each sentence with the pitch accent on the theme altered so that each of the four parameters appeared in each of its two hypothesised settings ('t'-like and 'r'-like), see Figure 1. 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 [3, chap.7].

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, as similar previous studies have shown speakers have a preference for the second version [4]. 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. This made a total of 256 dialogue pairs for each subject, which were presented in 16 blocks of 16 randomly-ordered pairs. Thirty subjects, staff and students at the University of Edinburgh, took part.

[3]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.

2.2.2. *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 and by the type of sentence. Using a 1 x 2 repeated-measures ANOVA there was a significant main effect of order, $F(1, 24) = 6.508,\ p = 0.018$; and place, $F(1, 24) = 4.617,\ p = 0.042$. The variables interacted, in theme-rheme order, when subjects heard the 'good' 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 ('good'-'bad':74.3%, 'bad'-'good':75.4%). Only Height caused subjects to significantly prefer the answer with that parameter in its 't' setting on the theme pitch accent. 73.4% of subjects chose the version of the answer in this category with lower pitch, significantly better than chance (using a 2 x 2 chi squared test $\chi^2 = 140.8,\ p < 0.01$).

The experiment seems to show that people, in production and perception, expect to hear a T accent in the theme position. It seems, however, that listeners only perceive the difference in rheme-theme ordering (taking the presentation order preference into account). The 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 prominent enough to signal the pitch accent on their own.

3. Interpretation of pitch height

The most robust finding, then, from these experiments is that the difference between theme and rheme accents is primarily signalled by pitch height. Both theme and rheme accents involve an H*, but this is lower for a theme accent than a rheme. There are several objections to this finding: impressionistically, listening to the stimuli for the perception experiment, an accent in the theme position with lowered pitch but all other parameters in the 'r'-setting did not 'sound' right, it feels as if there is something else going on.

Within ToBI this distinction can be coded at best indirectly. Themes could be coded as downstepped: either L+!H* or !H*, and rhemes not: L+H* or H*. But this is not helpful as one can in principle have either theme-rheme or rheme-theme order in a sentence, but you cannot have a pre-downstepped accent, e.g. a L+!H* L+H* sequence. Or you could say that the L target is triggering pitch lowering, however, since all themes and most rhemes in the production data seemed to be preceded by an L target, this seems a rather obtuse way to describe the phonetics of the two accents involved, and fails to deal with the apparent lack of distinction between theme and rheme accents at the beginning of a phrase.

Pierrehumbert & Hirschberg view the successive lowering of pitch peaks in an utterance as a narrowing of pitch span (as indeed is generally accepted), but see this as essentially a paralinguistic phenomenon. They say that this shows the hierarchical organisation of phrases within a discourse, with the pitch range lowering within each phrase and resetting at the beginning of a new phrase. This indeed, has been clearly shown to signal topic structure in monologues [2]. The relationship between pitch span differences and semantics has, however, to my knowledge, never been formalised. There is evidence that differences in pitch level can signal semantic distinctions beyond topic structure. In a series of experiments, Ladd & Morton [4] showed that listeners were willing to classify 'High' and 'Extra High' pitch accents as signalling a 'normal' vs 'emphatic' interpretation of the utterance, although it was unclear whether this was in fact a phonological distinction.

4. Three layers of pitch perception

This suggestion takes us to the reinterpretation of these experimental results proposed in this paper. Ladd [3, chap.7], in his discussion of how to deal with pitch span[4] in intonational phonology, claims that the phonetics of pitch range variation are affected in three distinct ways. Intrinsic effects have to do with the height and alignment of tonal targets within the pitch span, what ToBI is trying to model. Extrinsic effects are those affecting the pitch span itself. Global extrinsic effects act over large portions of speech, are either extralinguistic, e.g. male/female, or paralinguistic, e.g. bored/excited. Local extrinsic effects influence the pitch span of one phrase in relation to those around it, and clearly have some sort of semantic interpretation, such as topic structure. Ladd suggests that there is a metrical relation between the pitch spans of adjacent phrases.

Ladd's three-way division is very similar to Bolinger's [1], who proposes a four-way layering of pitch interpretation. Bolinger's second and third layer correspond to Ladd's local extrinsic effects layer, which he says conveys a number of partially grammaticised meanings relating to the affective meaning of the sentence. Specifically, he claims that a higher pitch span represents hearer-orientation, or lack of control by the speaker. Lower pitch spans, on the other hand, represent speaker-orientation, showing the speaker is assured, or speaking for his own benefit.

Since the target words in the previously reported experiments were all at the end of the phrase they were in, it is possible that the pitch height differences that were produced and perceived were in fact differences in the amount of pitch span narrowing in the two contexts. To test this, in a small pilot production study, we looked at six sentences such as (4) where it was possible to measure the phonetic properties of both the theme and the rheme in both positions in each phrase:

(3) Q: You're going to see Amanda tomorrow, right?				
A: No,	(I'm seeing	Amanda)	(on Monday),	
		theme	*rheme*	
	(I'll see)	(Norma)	(tomorrow).	
		rheme	*theme*	

The experimental set-up was very similar to the first production experiment, except this time a male speaker was used. This time, however, we found there was no significant difference in the peak height at the beginning of the phrase (whether it appeared first or second (Theme: H=189.3Hz; Rheme: H=191.9Hz, using a two-tailed paired t-test $P < 0.716$), whereas at the end of the phrase themes were significantly higher than rhemes (Theme: H=135.6Hz; Rheme: H=157.6Hz, using a two-tailed paired t-test $P < 0.001$). Further, we found

[4] which can be distinguished from overall pitch level, i.e. the starting point of pitch variation. But since the two co-vary and we are looking at data from only one speaker we will act as if it is enough to look only at pitch span, i.e. the distance between the top and bottom of a speaker's pitch range.

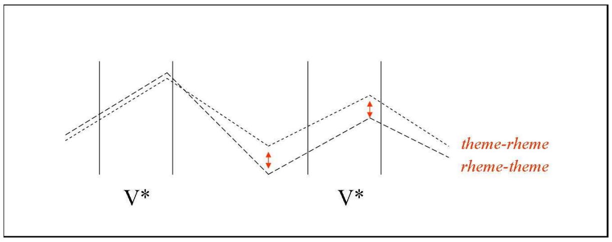

Figure 2: *Pitch range narrowing in second pilot production study*

the size of the fall after the first pitch accent to be significantly greater in the rheme-theme order than vice versa, also suggesting a Bolinger-style pitch span narrowing. This can be seen in Figure 2.

Taking Ladd's proposal that the relationship between the pitch spans of adjacent phrases is metrical, and Bolinger's intuition about the relationship between pitch span lowering and speaker orientation, informativeness and control, a new analysis emerges. The theme-rheme distinction is marked by the relative pitch spans of adjacent phrases. Given that pitch declines over the course of a topic (which each of our utterances is), theme-rheme ordering is shown by a low-high relation, and rheme-theme ordering is shown by a high-low relation. Themes and rhemes are still intonational phonological categories, however, they operate on the local extrinsic level of pitch.

If this view is taken, the phonetic effects found in each of our experiments fall out easily. The consistent finding of a fall following a rheme, but not a theme accent, is in fact a signalling of pitch span lowering. The alignment differences found in the first production study are a consequence of the phonetic space needed to reach an H target within a wider or narrower pitch span - in a narrower pitch span the rise can begin later than in a wider pitch span to reach the same target, hence the later alignment of the beginning of the rise. However, this alignment difference does not, in itself, signal the categorical difference, which is why it failed to have an effect when manipulated independently in the perception study. Finally, if we take it that the assignment of metrical structure works left-to-right, it makes sense that there is no difference in the pitch height of theme and rheme accents at the beginning of an utterance - as found in both the second production study and the perception study, where subjects only consistently judged theme accents to be more appropriate in rheme-theme order.

Given this analysis, the need for two phonological categories, L+H* and H*, becomes less clear. On phonetic grounds, there seems little reason not to either adopt the suggestion in Ladd & Schepman [5] that all accents preceded by a clear low should be reclassified as L+H*, leaving H* to describe purely high accents, or collapse the categories entirely. On phonological grounds, in the examples used in the present study, these accents seem to mark one semantic category, *focus*. However, further production and perception studies which control for pitch scaling effects may show that these categories in fact mark other semantic distinctions, such as speaker commitment, suggested in [6].

5. Conclusion

In this paper I have presented evidence for a richer phonological interpretation of intonation in English. I have reported studies and given new experimental results which seem to show that phonetic intonational cues, specifically pitch variation and segmental alignment, are the result of the interaction of three distinct levels of intonational effects, at least two of which (intrinsic and local extrinsic) have a definable semantic interpretation.

This conclusion resulted from an investigation of two supposed intonational phonological categories, theme and rheme pitch accents. The characterisation of these in terms of intrinsic effects - i.e. ToBI categories, L+H* and H* - has proved elusive. However, the phonetic differences between the two accents found in this series of production and perception experiments can be nicely explained in terms of relative pitch span variation, a local extrinsic effect. If a metrical view of the relative pitch spans of adjacent phrases is taken, theme-rheme ordering is indicated by a low-high relation, and rheme-theme ordering by a high-low relation.

It is clear that if these two supposed levels of intonational effects do, in fact, directly affect the semantic interpretation of utterances, then studies into the phonetic dimensions of intonational categories need to take careful account of the interaction of the phonetic effects in the two levels. There is definitely room for much investigation into the formal semantics of the local extrinsic level of intonation.

Additional information about the experiments reported in this paper, including sound files, may be found at the following website: *http://www.cstr.ed.ac.uk/ s0199920/research.html*

6. References

[1] Bolinger, D., 1970. Relative Height. In *Prosodic Feature Analysis*, P. Leon (ed.). Montreal:Marcel Didier, 109-127.

[2] Grosz, B.; Hirschberg, J., 1992. Some intonational characteristics of discourse structure. In *Proceedings of the International Conference on Spoken Language Processing*, Banff, Canada, 429-432.

[3] Ladd, D.R., 1996. *Intonational Phonology*. UK: Cambridge University Press.

[4] Ladd, D.R.; Morton, R. 1997. The perception of intonational emphasis: continuous or categorical?. *Journal of Phonetics*, 25, 313-342.

[5] Ladd, D.R.; Schepman, A., 2003. 'Sagging transitions' between high pitch accents in English: experimental evidence. *Journal of Phonetics*, 31, 1, 81-112.

[6] Pierrehumbert, J.; Hirschberg, J., 1990. The Meaning of Intonational Contours in the Interpretation of Discourse. In *Intentions in Communication*, P. Cohen, J. Morgan & M. Pollack (eds.). Cambridge, MA: MIT Press, 271-311.

[7] Pierrehumbert, J.; Steele, S., 1989. Categories of Tonal Alignment in English. *Phonetica*, 46, 181-196.

[8] Silverman, K.;Beckman, M.;Pitrelli, J.; Bailly, G.; Aubergé, V., 1997. Phonetic representation for intonation. In *Progress in Speech Synthesis*, J.Ph. van Santen (ed.). New York: Springer, 435-441.

[9] Steedman, M., 2000. Information Structure and the Syntax-Phonology Interface. *Linguistic Inquiry*, 31, 4, 649-689.