Do H*L and L*H accents have similar target positions?

Toni Rietveld & Joop Kerkhoff

Radboud University Nijmegen

Abstract

Targets of pitch accents have generally been defined on the basis of production data. Obviously, most of the research on tonal targets has focussed on peaks associated with H*L accents as these are very well identifiable, in contrast to the lows associated with L*. In the perception experiments reported here it is made clear that, in Dutch, the conventional target positions of L* and H* (being the F_0-valley and the F_0-peak respectively) do not coincide. The valley associated with L* which gives rise to the perception of a sentence accent occurs much earlier in a syllable than the peak associated with H*. No significant effect was found for the position of the accent: nuclear or prenuclear. Confirmation was found for the existence of the phonological contrast between H* H% and L*H H% in Dutch.

1 Introduction

In the autosegmental tradition F_0-targets have been generally defined on the basis of production data; they are "considered to be identifiable points in the F_0-contour which are aligned with the segmental string in extremely consistent ways" (Ladd, 1996:67). Knowledge about the alignment of targets in production does not necessarily mean that we know which part of a pitch contour associated with a pitch accent is used by the listener to *locate* the accent. As a matter of fact, quite a number of 'turning points' are candidates. For an H*L accent obvious candidates are the start of the rise and the position of the F_0-peak. For L*H-accents the relevant turning points are less evident. The fall associated with a %L L*H-accent is quite small, and is sometimes hardly distinguishable from the low onset. Obviously, most of the research on tonal targets has focussed on peaks associated with H*L accents as these are very well identifiable, in contrast to the lows associated with L*. Arvatani, Ladd & Mennen (1998) presented one of the few investigations on the alignment of L*-accents (for Greek); the F_0-minimum was found to be aligned in a consistent way: 5 ms before the onset of the accented syllable. In many respects the pitch accent L*H takes a somewhat special position in intonation: it seems to be acquired in a rather late stage of speech development (Bolinger, 1989). To 'novices' in the study of intonation it looks often counterintuitive that focus can be signalled by a low target: in pilot studies we find again and again that naive subjects tend to assign a pitch accent to the syllable in which the final F_0-point of the H of L*H is located.

Doubts about the validity of the assumed equivalence of high and low target F_0-values for H*L and L*H might come from the difference in tonal prominence between reaching the 'peak' and the 'valley', respectively, of the two pitch accents at issue. We think, therefore, that the concept of 'target' should also be considered from a perceptual point of view; the comparison of the perceptual processing of the pitch accents H*L and L*H might shed more light on the correspondence between productive and perceptual targets. Furthermore, it is difficult to determine which tonal event in a nuclear L*H-accent gives rise to the perceived accent: the early valley associated with nuclear L*, or the high F_0-value associated with the following H, which is (nearly) in the same syllable (σ_1), see Figure 1.

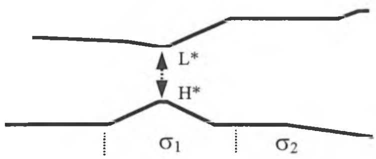

Figure 1. Possible targets of L*H and H*L.

House (1989) developed a model according to which tonal movements through areas of spectral change will be optimally categorized as levels, instead of movements: a falling movement as Low, and a rising movement as High. House assumes that at vowel onset the perceptual mechanism is maximally loaded with the task of resolving spectral information; thus its capacity to resolve F_0-movements is decreased. This would mean that a falling F_0-movement extending over the CV-boundary leads to the perception of a Low in the vowel, and a rise extending over the CV-boundary to the perception of a High. Thus, in a nuclear L*H-accent, a steep rise extending over the VC-boundary, might be perceived as a 'high' target on the vowel following the V associated with L*. As a consequence this vowel might be perceived as carrying the accent, because in the competition of high and low targets it is likely to dominate (Wales & Taylor, 1987).

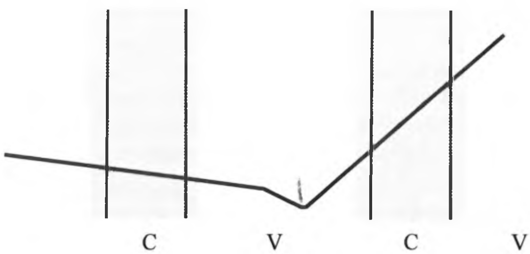

Figure 2. According to House (1989) tonal movements through areas of spectral change (CV and VC) will be categorized as levels, instead of movements: a falling movement as L, a rising movement as H.

In the experiments to be reported here we assessed to what extent the perceptual effects of H*L and L*H accents are similar when the initial boundary tone is %L; this initial boundary tone was chosen as it yields the smallest fall associated with L*. Perceptual effects are restricted to the perceived position of a pitch accent in the syllable as a function of specific tonal events, like reaching a target F_0-value (F_0-maximum for H*L and F_0-minimum for L*H). Furthermore we want to show that the perceptual effects of the pitch accent L*H differ as a function of the location of this accent in the sentence. We expect nuclear L*H-accents, with quite steep F_0-movements to the target of H, to yield different perceptual reactions to shifts in the position of the L*-target from non-final L*H-accents, in which the linking with the following accent leads to smoothly rising F_0-movements (cf. Gussenhoven & Rietveld, 2000). This specific position also creates the possibility to assess to what extent listeners use global information, based on the perception of a whole pitch contour, or use specific tonal events, like a pitch valley, associated with L*. The question whether subjects use global information to locate pitch accents can be answered by deleting the valley associated with L* in %L L*H H%-sequences, and presenting the resulting pitch contour to listeners.

There is another, less general, because language-specific question which needs confirming evidence. Both in the descriptions of Dutch intonation of 't Hart, Collier & Cohen (1990) and in that of Gussenhoven & Rietveld (1991) the contrast between H* H% and L*H H% was not recognized; the latter equated Pierrehumbert's (1980) H* H% with an L*H H% contour pronounced with wide span in a higher pitch register. Gussenhoven & Rietveld (2000) concluded on the basis of scores on gradient paralinguistic attributes (like the degree of 'SURPRISE' conveyed by these contours) that this contrast does exist in Dutch. Establishing differences in the target positions of both contours would be regarded as final confirmation.

To summarize, we aim at finding answers to the following questions:

Q1: Do the perceptual target spaces of H*L and L*H have the same alignment?

Q2: Is the perceptual target of L*H the same in nuclear and prenuclear positions?

Q3: Do listeners use global F_0-information when specific local F_0-cues are missing?

Q4: Does the contrast between H* H% and L* H H% contours exist in Dutch?

2 Method

2.1 Speech materials

Two source utterances were synthesized with the Nijmegen/MBROLA diphone synthesis system (male voice; sampling frequency 16 kHz): *hij wil mo mo verlaten* ('he wants to leave mo mo') and *hij wil mo mo voor een tijdje verlaten* ('he wants to leave mo mo for some time') with the quasi nonsense syllables *mo1* (the first *mo* in the utterance) and *mo2* (the second *mo*), which together suggest a name,). No durational, intensity or spectral cues were assigned to either *mo1* or *mo2* to signal the location of the accent.

Six experimental contours were assigned to the source utterances, the first three to the shorter, the latter three to the longer utterance:

(1)	%L H*L L%	('pointed hat')
(2)	%L H* H%	(equivalent to a %L L*H H%-contour without valley associated with L*)
(3)	%L L*H H%	(valley of default 12 Hz).
(4)	%L H*LH !H*LH H%	(prenuclear accent on '*la*' of *verlaten* - 'to leave' -)
(5)	%L L*H !H*LH H%	(prenuclear accent on '*la*' of *verlaten*: linked contour)
(6)	%L L_0*H !H*LH H%	(as (5), but without a valley associated with L*, L_0 stands for a deleted L-target)

In Figure 3 the six contours are depicted.

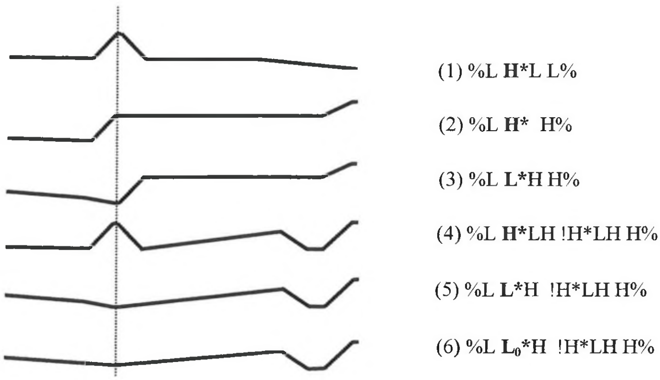

Figure 3. The alignments of the six contours used in the perception experiments. The dotted line marks the conventional targets of the pitch accent, at issue.

These contours were presented to subjects, with different (shifted) temporal locations of the targets; the step size was 20 ms, see Figure 4.

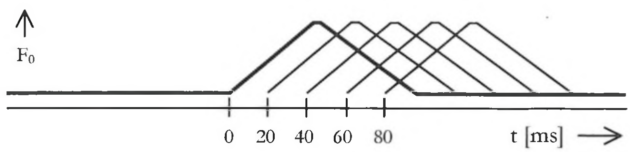

Figure 4. Shifts of the pitch configurations in steps of 20 ms: example for H*L .

The task of the subjects was to indicate whether the accent is on the first or on the second *mo*.

Below we give a detailed account of the information the responses to the six contours (see Figure 3) can give us:

(1) The first contour (with H*L) can be seen as an anchor contour, with a generally accepted target, the F_0-peak.

(2) The second contour (with H* H%) has nearly the same make-up as the L*H-contours, but is not realised with a valley. This contour enables us to find out two things:

a. whether the same F_0-movement, associated with reaching H*, will be interpreted differently as a function of the tonal material which follows; this would mean that listeners try to give an interpretation to the whole contour, and do not (only) use specific target values, and

b. to which extent this contour is processed differently from similar contours in which the valley is realized (contour 3). If not, than we have to assume that the valley itself is not a crucial part of this contour. The comparison with contour 3 (%L L*H H%) enables us to assess whether these contours constitute a phonological contrast.

(3) Contour 3 is a default realisation of the L*H-accent; this contour provides information on the question whether the hypothesis is correct that the highest and lowest F_0-value of H*L and L*H-accents, respectively, should be seen as the targets of these accents.

(4) Contour 4 is a realisation of the %L H*LH !H*LH H% accent; it is the two-accented stimulus to be used as 'standard' to contours (5) and (6) in which the prenuclear accent is L*H.

(5) Contour 5 is a default realisation of the L*H accent in prenuclear position (linked with H*LH on *la* of *verlaten*). Thus, a smooth pitch rise to the following accent can be realized. If shifting L*-targets yields accent assignments that differ from those obtained in the single L*H-accents of contour 3, we will have extra evidence that L*H is not a unit, which functions independently of tonal context, like – as we predict – in the case for H*L

(6) Contour 6 is a non-default realization of the L*H-accent, in that the valley associated with L* is deleted (L_0*). If subjects still exhibit the same perceptual patterns as with contour 5, we will have evidence that they use global, contour-based information to assign pitch accents, and not only specific tonal events, strongly associated with the 'accented'syllable.

2.1.1 Physical characteristics of the contours

The duration of the movement towards the target of H* is 100 ms (93 Hz to 147 Hz), the duration of the F_0-movement towards the valley is also 100 ms; the valleys in the %L L*H H% accents are 12 Hz below the preceding last F_0 value of the onset. The %L-onsets start at 104 Hz and end at 93 Hz. The duration of the H% movement is 120 ms, it covers 33 Hz with 180 Hz as endpoint. The F_0-maxima of H* and H in L*H are the same: 147 Hz.

The respective durations of [m] and [o:] are the same in both syllables *mo*. The positions of the conventional targets for L*H and H*L (start of the valley and position of the maximum F_0-value) are shifted in a variable number of steps (between 10 and 15) of 20 ms from different starting points.

In Table 1 we give the start and end of the segments in the two syllables *mo*, and positions of the turning points of the five experimental contours.

Table 1. Start and end of segments in the two syllables *mo*, and positions of turning points in ms (beginnings of conventional targets: maximum F_0 for H* and minimum F_0 for L*, see fig. 3) of the six experimental contours, starting from the 'left'.

	m	o	m	o
Start/end of segment	425-485	485-630	630-690	690-835
Contour				
(1) %L **H***L L%		560		740
(2) %L **H*** H%		600		780
(3) %L **L***H H%		500	680	
(4) %L **H***LH !H*LH H%		520		760
(5) %L **L***H !H*LH H%		520		760
(6) %L **L_0***H !H*LH H%		520		760

The total duration of the short sentences (with contours 1, 2 and 3) was 1510 ms, that of the long sentences (with contours 4, 5 and 6) was 2160 ms.

2.2 Task and procedure

Subjects had to tell whether they heard the name *MOmo* (*mo1* with accent ('stress') on the first syllable: 'a') or *moMO* (*mo2* with accent on the second syllable: 'b').

Two methods of presentation were used, (A) that of minimal changes, in which the stimuli of a specific contour are presented in a specific order, either with F_0-targets shifting to the right ('ascending') or to the left ('descending'), and (B) that of random presentation, in which the stimuli were presented in random order. The stimuli were presented to four groups of 10 subjects, to whom different methods of presentation were assigned (either method A or method B, with 20 subjects each) and different orders of stimuli. Each utterance is followed by an response interval of 3 seconds.

For *presentation method A*, each block of stimuli was preceded by two 'anchor stimuli': a default 'a' and a default 'b' contour with unambiguous accents on the first and the second *mo* syllable, respectively. The anchor stimuli always corresponded with the particular contour to be presented in that group. In the instruction for the listeners it was told that the number of stimuli per block could vary between 10 to 15; thus we tried to avoid expectations about the stimulus in which the change from *a* to *b* or vice-versa had to be expected. Accordingly the score form contained 15 answer positions.

For *presentation method B*, the whole experiment was preceded by default realisations of all contours to be presented, with 'clear' accents on syllables 'a' or 'b' respectively. Each presentation cost about 10 minutes. Stimuli were presented over earphones; subjects were free to determine the start of the presentation of each group of stimuli.

3 Results

Not all our data fit the psychometric curve; in the data obtained with method A the transition from 'a' to 'b' judgements took place in a very categorical way. That is why it was decided to carry out analysis of variance of the randomized block type on the data obtained with method A, followed by post-hoc comparisons. The six contours make up the within-subject factor 'treatment'. For each subject the time (in ms) is given at which the target of a particular contour gives rise to a 'b' judgement. Subsequent post-hoc comparisons show which contours have different or similar 'perceptual target values'. For method B the application of the conventional psychometric curve fitting procedure was not problematic for most contours, because the transitions from 'a' to 'b' judgements were much smoother, within subjects and pooled over subjects.

3.1 Results obtained with method A

In Table 2b we present the mean transition values, pooled over 20 subjects; for clarity's sake we reproduce part of Table 1 in Table 2a, with the starts and ends of the segments at issue.

Table 2a. Locations of starts and ends of the segments of *mo mo,* in ms.

m	o	m	o
425...485	485...630	630...690	690...835

Table 2b. Mean locations of transition from ‘a’ to ‘b’ responses, in ms, obtained with method A. Each value is based on 20 observations. Relevant targets (L* or H*) printed in bold.

Contour	Position of transition
(1) %L **H***L L%	685
(2) %L **H*** H%	701
(3) %L **L***H H%	628
(4) %L **H***LH !H*LH H%	704
(5) %L **L***H !H*LH H%	644
(6) %L **L_0***H !H*LH H%	655 (?)

The perceptual effects of shifting the target of **H*** in %L **H***L L% (contour 1) and %L **H*** H% (contour 2) are nearly the same. In %L **H***L L% the location of the target at 685 ms leads to the perception of an accent shift from *mo*1 to *mo*2, and in %L **H*** H% at 701 ms, a difference within the step size of 20 ms. This is the location of the onset of [o:] in *mo*2.

For **L*** in %L **L***H H% (contour 3) the position is clearly different: the perception of an accent shift from *mo*1 to *mo*2 takes place at 628 ms, much earlier than with H* (at 701 ms). This is apparently the position where H following L* reaches its maximum value (at 728 ms). This means that the perceptual anchors of H* and L* are not the same in final accents.

It is not surprising – at least from a perspective which focuses on the use listeners may make of physical cues – that the perceptual effect of shifting the target of **H*** in %L **H***LH !H*LH H% (contour 4) is about the same as in both %L **H***L L% (contour 1) and %L **H*** H% (contour 2): the perceptual shift took place around 700 ms. For **L*** in %L **L***H !H*LH H% (contour 5) the situation is not so clear. In the series in which the target was shifted from left to right, the target position of L* at which *mo*2 was perceived as accented was 670 ms, whereas in the series with right to left shifting, the location was 620 ms, a difference of about 2 stimulus shifts. They are not so different from the position of **L*** in %L **L***H H% (contour 3), while both are different from those observed with an H*-accent.

Very unclear are the perceptual effects of contour (6) %L **L_0***H !H*LH H% in which **L*** is not marked by a valley. Nine of the 20 subjects did not hear any transition, two others started a series with a perceptual localisation opposite to the direction of the shifts. The nine remaining subjects located the shift at 655 ms.

An ANOVA was carried out on the location of the transition for each contour, apart from contour (6): %L L_0*H !H*LH H%, for which a clear transition could not be established; the (fixed) factor ‘contour’ was significant: $F(4,56)=24.69$, $p<.01$. Post-hoc comparisons (Tukey’s HSD procedure; $\alpha=.05$) yielded the following homogenous subsets (mean values of transition locations given in ms):

Table 3. Homogeneous subsets of contours (time position of accent transition, in ms).

Contour	subset 1	subset 2
(3) %L L*H H%	628	
(5) %L L*H !H*LH H%	644	
(1) %L H*L L%		685
(2) %L H* H%		701
(4) %L H*LH !H*LH H%		704

Table 3 reveals that there is a clear distinction between alignments associated with L* and H*-accents respectively. The existence of a contrast between %L L*H H% (3) and %L H* H% (2) is confirmed again.

3.2 Results obtained with method B

The results obtained with method B were processed in the conventional way: the percentages of judgements obtained by pooling over subjects were probit-transformed and the stimulus value which coincided with $z=0$ was regarded as the Point of Subjective Equality (PSE). None of the associated χ^2 values used to test the fit were significant at the .05 level, which reflects goodness-of-fit, except for the results of one contour: %L H*LH !H*LH H% (contour 4). For this contour $\chi^2(7)=30.86$, $p<.01$. The reason for this poor fit was the steep transition from 'a' to 'b' judgement. As estimate of the PSE we took the position at which the majority of the subjects had given a 'b' judgement.

For contour (6) %L L_0*H !H*LH H% (without a valley for L*) neither a conventional sigmoid was observed (its direction was even nearly reversed, and the percentages of 'b' responses varied around 60%), nor a clear transition in the perceived location of the accent. In fact, the situation was analogous to the one found with method A.

In Table 4 we present the mean transition values obtained for the six experimental contours, obtained with the two methods. Each mean value is based on 20 observations.

Table 4. Mean locations of transition from 'a' to 'b' responses in ms, obtained with methods A (series) and B (random order). Each value is based on 20 observations.

Contour	Transition (A)	Transition (B)
(1) %L H*L L%	685	671
(2) %L H* H%	701	704
(3) %L L*H H%	628	631
(4) %L H*LH !H*LH H%	704	700
(5) %L L*H !H*LH H%	644	640
(6) %L L_0*H !H*LH H%	655 (?)	737 (?)

Table 4 shows similar transition points for all but one contour (number 6); with both methods A and B we see nearly equal points of transition for accents with H* and with L* respectively.

In Figure 5 we summarise our findings by showing the alignments of the contours which mark the positions of Perceptual Subjective Equality; the alignment of contour %L L_0*H !H*LH H% (6) is not given, as we could not determine a clear PSE for it. (The slight differences of tonal height among the H*-contours only serve pictorial clarity).

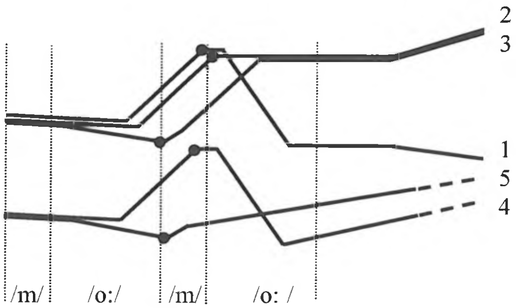

Figure 5. Positions of the conventional targets (for H* the beginning of the F_0-maximum and for L* the beginning of the valley, • in the figure) that give rise to the perception of an accent on the second 'mo' in the series. The numbers refer to the contours given in Table 4.

The results summarised in Table 3 and Figure 5 make four things very clear. It is not correct to equate the conventional alignments of L* and H*-accents, viz. the location of the start of the valley and the location of the F_0 maximum. The small valley associated with L* is an important cue; we do not know, however, whether this cue plays a local role, or whether it triggers the use of the rest of the contour to locate the accent. No difference was found between the perceptual processing of nuclear and prenuclear pitch accents. The H* H% and L*H H% contours constitute a phonological contrast, as the timing of the associated target values is completely and significantly different.

4 Conclusion

The perceptual shift from an accent on the first syllable of the sequence *mo mo* to the second takes place earlier with L*H than with H*L accents, at least when the targets are expressed in the positions of the conventional targets of these accents. The difference, pooled over the variants of the two accent types amounts to about 50 ms, and confirms the phonological contrast between H* H% and L*H H% contours in Dutch, earlier established on the basis of paralinguistic scale judgements.

Whereas the presence of a valley at the onset of the sonorant consonant of the second syllable *mo* already gives rise to the perception of a sentence accent, the target of H* needs to be located at the onset of the vowel of that syllable, to be prominence lending. One of our hypotheses was that the H following the L* in nuclear accents might compete with the F_0-valley associated with L*, analogously to the way in which H-targets can compete with L*-targets in the degree of perceived prominence (cf. Gussenhoven & Rietveld, 2000). The introduction of L* in prenuclear position in contour (5) %L L*H !H*LH H% gave us the opportunity to test this hypothesis. Indeed, the target of prenuclear L* (with the following slowly rising H) gave rise to a perceived accent on the second syllable associated with a later target; the difference is 16 ms, but it is not statistically significant at the 5% level. Thus, until other evidence becomes available, we have to assume that the targets of L* in both prenuclear and nuclear positions behave similarly, but they should be accounted for by different explanations. For nuclear L*H, House's theory (1989) might throw light on the results. He assumes that at vowel onset the perceptual mechanism is maximally loaded due to the spectral instability at vowel onset; consequently its capacity to resolve F_0-movements is decreased. A rise extending over the onset of the vowel following the position of the low target might be 'categorized' as a high target in that vowel, and consequently lead to the perception of a H*-accent on that vowel. If the steep rise leads to a relatively high F_0 value in the same vowel in which the F_0 valley occurs, it might even be perceived as delayed H*. Thus, an L*H accent might go through different perceived accents, L* and H*, depending on the location of both L*-targets and targets and movements associated with the following H.

However, this explanation does not account for the behaviour of L*H in prenuclear position. First of all, the presence/absence of the small valley in contours (5) %L L*H !H*LH H% and (6) %L L_0*H !H*LH H% appears to be crucial for the perception of a prenuclear L*H accent. At first sight, one might be inclined to think that the hypothesis of 'global perception' of contours must be rejected, as a local cue – the small valley of 12 Hz with a duration of only 15 ms, preceded by a downwards slope from 104 to 93 Hz over 100 ms – plays a role in the perception of the associated accent. We think however that it is not wise to dismiss the concept of 'global perception' altogether. In Dutch, %L (–) H*L is not a possible contour with a slowly rising pitch, starting somewhere at (–) in the contour before H*, unless the start of the slope is associated with either the realisation of an H*L or an L*H-accent. This constraint on possible contours might make listeners search for F_0-cues which mark the presence of an L*H-accent, as an H*L-accent is clearly lacking. Thus, 'global perception' might interact with the presence/absence of local cues, and give rise to the perception of accents.

References

Arvatani, A., Ladd, D.R., Mennen, I. (1998). Stability of tonal alignment: The case of Greek prenuclear accents. *Journal of Phonetics, 26*, 3-25.

Bolinger, D. (1989). *Intonation and its uses: melody in grammar and discourse.* London: Edward Arnold.

Gussenhoven, C.H.M. & Rietveld, A.C.M. (1991). An experimental evaluation of two nuclear tone taxonomies. *Linguistics, 29,* 423-449.

Gussenhoven, C. & Rietveld, T. (2000). The behaviour of H* and L* under variations in pitch range in Dutch rising contours. *Language and Speech, 43*, 183-203.

't Hart, J., Collier, R., Cohen, A. (1990). *A perceptual study of intonation.* Cambridge: Cambridge University Press.

House, D. (1989). Perceptual constraints and tonal features. *Working Papers, Department of Linguistics, Lund University, 35*: 113-120.

Ladd, D.R. (1996). *Intonational phonology*. Cambridge: Cambridge University Press.

Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation.* PhD. thesis, MIT.

Wales, R. & Taylor, S. (1987). Intonation cues to questions and statements: how are they perceived? *Language and Speech*, *30*, 199-210.