Speech Prosody 2002
Aix-en-Provence, France
April 11-13, 2002

The Temporal Alignment of L*H Accents

Toni Rietveld & Joop Kerkhoff

University of Nijmegen, Department of Language and Speech,
P.O. Box 9103, 6500 HD Nijmegen
The Netherlands
A.Rietveld@let.kun.nl

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 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" [5]. For an H*L accent obvious candidates are the start of the rise and the position where the F_0-peak occurs. 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. [2] showed that the degree of perceived "prominence" of a %L L*H H%-contour depends on the absolute depth of the valley associated with L* or the height of H%, as a function of the distance between L* and the endpoint of H%. Their results show that, unlike H*L, the 'non-starred' tone segment plays a role in the perception and processing of the L*H-accent.

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*. Doubts about the validity of the assumed equivalence of high and low target F_0-values for H*L and L*H respectively 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. In many respects the pitch accent L*H takes a somewhat special position in intonation: it seems to be acquired in quite a late stage of speech development [1]. Furthermore, we cannot determine which tonal event in a nuclear L*H-accent is crucial: the valley associated with L* or the following high F_0-value. [3] 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. [3] 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 conventionally 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.

In the experiments to be reported here we assessed to which extent the perceptual effects of H*L and L*H accents are similar when the boundary tone is %L. 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 prenuclear L*H-accents, in which the linking with the following accent leads to smoothly rising F_0-movements. This specific position also creates the possibility to assess to which extent listeners use global information, based on the perception of a whole pitch contour, or use specific tonal events, like a pitch valley. The question whether listeners 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. [2] concluded on the basis of scores on gradient paralinguistic attributes (like the degree of 'SURPRISE' conveyed by these contours) that the contrast between H*H% and L*HH% contours does exist in Dutch. A confirmation by the establishment of differences in the target positions of both contours should be regarded as the final one.

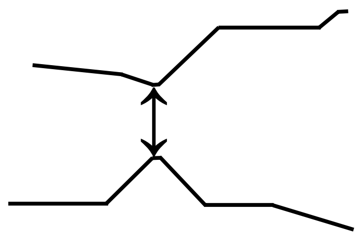

Figure 1: The default targets of L*H and H*L in the Nijmegen Speech Synthesis System.

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: *hij wil MO MO verlaten* (Eng.: 'he wants to leave MO MO'), with one (perceived) accent on either the first or the second MO and *hij wil MO MO voor een tijdje verLAten* (Eng.: 'he wants to leave MO MO for some time') with, apart from possible perceived accents on either the first or the second MO, a nuclear H*LH accent on '*la*' of *verlaten.* The two quasi nonsense syllables: MO1 (the first MO in the utterance) and MO2 (the second MO), which together suggest a name, were synthesized with the Nijmegen-/MBROLA diphone synthesis system (male voice; sampling frequency 16 kHz). 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*+L !H*LH H% (nuclear accent on '*la*' of *verlaten*)
5. %L L*H !H*LH H% (nuclear accent on '*la*' of *verlaten*: linked contour)
6. %L L_0*H !H*LH H% (as (5), but without a valley associated with L*)

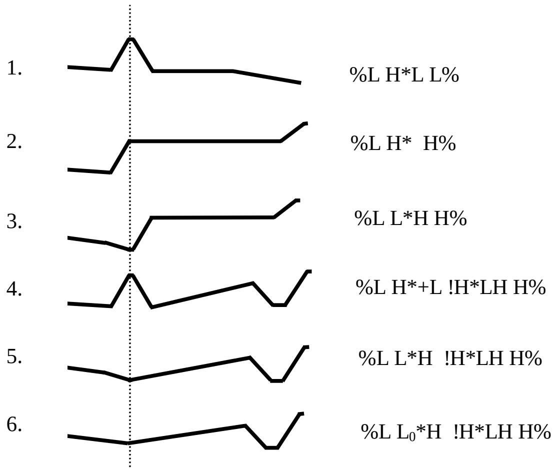

Figure 2: The alignments of the six contours used in the perception experiments.

These contours were presented to subjects, with different (shifted) temporal locations of the targets; the stepsize was 20 ms, see figure 3. The task of the subjects was to indicate whether the accent is on the first or on the second *mo*.

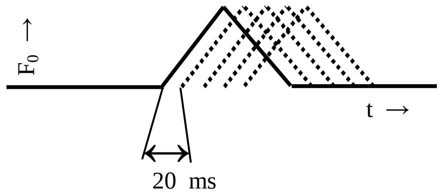

Figure 3: Shifts of the pitch configurations in steps of 20 ms: example for H*L .

Below we give a detailed account of the information the responses to the six contours can give us:
Contour 1 (with H*L) can be seen as an anchor contour, with a generally accepted target, the maximum F_0-value, associated with H*.
Contour 2 (with H* H%) has nearly the same make-up as the L*H-contours, but is not realized 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 *(contour 1 vs. 2)*; if that is the case, than we must assume that some post-processing takes place in the interpretation of F_0-contours, 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. Another, important topic is that the comparison with contour 3 (%L L*H H%) enables us to assess whether these contours constitute a phonological contrast.
Contour 3 is a default realization of the L*H-accent; the presentation of this contour to subjects will give information on the question whether the hypothesis 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 is correct.
Contour 4 is a realization of the %L H*+L !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.
Contour 5 is a default realization 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 unity, which functions independently of tonal context, like – as we predict – is the case for H*L.
Contour 6 is a non-default realization of the L*H-accent, in that the valley associated with L* is deleted. 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.

2.1.1. Physical characteristics of the contours

The duration of the movement towards the target of H* is 100 ms, (from 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 %L-onset. The %L-onsets start at 104 Hz and end at 93 Hz. The duration of the movement associated with H% is 120 ms; the movement covers 33 Hz, and its endpoint is 180 Hz. The F_0-maximum in the contours with H* is the same as the F_0 associated with H in the L*H-contours: 147 Hz.

The respective durations of [m] and [oː] in both syllables *mo* are the same. The positions of the conventional targets for L*H and H*L (being the position where the valley is reached and the position where the maximum F_0-value is reached respectively) are shifted in 10 to 15 steps 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 six experimental contours.

Table 1: Start and end of segments in the two syllables 'mo', and positions of turning points of the six experimental contours in ms, starting from the 'left'.

	m	oː	m	oː
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*+L !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) 2160 ms.

2.2. Procedure

2.2.1. Task

Subjects had to tell whether they heard the name "*MO mo*" (with accent ("stress") on the first syllable: 'a') or "*mo MO*" (with accent on the second syllable: 'b').

2.2.2. Method of presentation

Two methods of presentation were used, that of minimal changes (A) and that of random presentation (B).
Method A: the method 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'). Six experimental contours and one filler contour were included in the experiment; the stimuli were presented in different blocks with different ascending and descending orders.
Method B: the method of random presentation, in which the stimuli were presented in random order. There were four random orders. The stimuli were presented to four groups of 10 subjects, to whom different methods of presentation were assigned (either series of subsequent shifts of the same kind of contour, or random presentations of stimuli, with 20 subjects each) and different orders of stimuli. Each utterance of about 1.5 or 2.2 seconds is followed by a response interval of 3 seconds.
For *method A*, each group of stimuli was preceded by two "anchor stimuli": a default 'a' and a default 'b' contour; 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 group could vary between 10 to 15.
For *method B*, the whole experiment was preceded by default realizations of all contours to be presented, with 'clear' accents on syllables 'a' or 'b' respectively.

Each presentation cost 10 minutes. Stimuli were presented over earphones.

3. Results

3.1. Introduction

Analysis of variance of the randomized block type was carried out on the data obtained with method A, followed by post-hoc comparisons. The six contours make up the within-subject factor 'contour'. For each subject the time (in ms) was 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'.

The analysis of variance 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: $F_{4,56} = 24.69$, $p < 0.01$. Post-hoc comparisons (Tukey's HSD procedure; alpha-level set at 5%) carried out on the data yielded the following homogenous subsets (mean values of transition locations given in ms):

Table 2: Homogeneous subsets of contours with locations of transitions of perceived accents on MO1 to MO2 in ms.

Contours	**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*+L !H*LH H%		704

Table 2 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.

For method B the application of the conventional psychometric curve fitting procedure was not problematic for most contours, as the transitions from 'a' to 'b' judgements was, within subjects and pooled over subjects, much smoother. The results obtained with method B were processed in the conventional way: the percentages of judgements obtained by pooling over subjects were probit-transformed into *z*-values. The stimulus value which coincided with $z = 0$ was regarded as the Point of Stimulus Equality (PSE). The associated χ^2 values used to test the fit were all not significant at the 5% level, which reflects goodness-of-fit, but for the results of one contour: %L H*+L !H*LH H% (4). For this contour $\chi^2_7 = 30.86$, $p < 0.01$. The reason for this bad 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 locations of the transitions from a perceived accent on the first to the second word.
In Figure 4 we summarize 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.

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*+L !H*LH H%	704	700
5) %L L*H !H*LH H%	644	640
6) %L L_0*H !H*LH H%	655 (**?**)	737 (**?**)

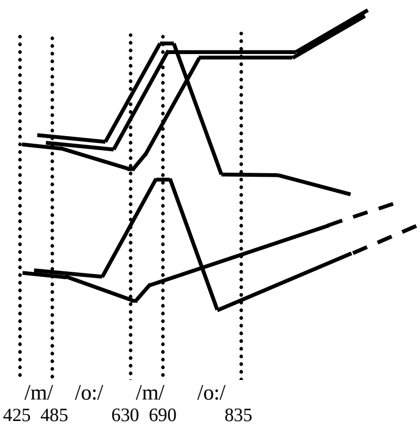

Figure 4: Positions of the conventional targets (for H* the beginning of the F_0-maximum and for L* the beginning of the valley) which give rise to the perception of an accent on the second 'mo' in the series (A).

The results summarized in Table 4 and Figure 4 make four things very clear:
a) 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;
b) The small valley associated with L* is an important cue; we do not know, however, whether this cue plays an isolated role, or whether it triggers the use of the rest of the contour to locate the accent.
c) No difference was found between the perceptual processing of nuclear and prenuclear pitch accents;
d) 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 conventional targets of H*L and L*H accents (the F_0- maximum and minimum respectively) are not similar. 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 contours in Dutch, earlier established on the basis of paralinguistic scale judgements.

Whereas the presence of a valley at the onset of the consonant of the second syllable *mo* of the sequence *MO 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. It is not very simple to explain the difference between the perceptual effects of the two traditional targets. One of our hypotheses was that the H following the L* in nuclear accents might compete with the F_0-valley associated with L*, analogous 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 is not 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 their positions should be accounted for by different explanations. For nuclear L*H House’s theory [3] might throw light on the results, but, unfortunately, it 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 short and small valley of 12 Hz, 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’ interacts with the presence/absence of local cues, and might still give rise to the perception of accents.

5. References

[1] Bolinger, D., 1989. *Intonation and its uses: melody in grammar and discourse*. London: Edward Arnold.
[2] Gussenhoven, C.; Rietveld, T. , 2000. The behaviour of H* and L* under variations in pitch range in Dutch rising contours. *Language and Speech* 43(2), 183-203.
[3] House, D., 1989. Perceptual constraints and tonal features. *Working Papers, Department of Linguistics*, Lund University 35: 113-120.
[4] Kerkhoff, J.J.P.; Rietveld, T., 1995. The generation of prosody in the Nijmegen Rule Oriented Speech synthesis system. In *ESCA 4th European Conference on Speech Communication and Technology,* Pardo, J.M.; Enriquez, E.; Ortega, J.; Ferreiros, J.; Macias, J.; Valverde, F.J. (eds.).
[5] Ladd, D.R., 1996. *Intonational phonology*. Cambridge: Cambridge University Press.

Acknowledgements

We would like to thank Carlos Gussenhoven for many useful comments regarding the design of the experiment.