

Explaining Attitudinal Ratings of Dutch Rising Contours: Morphological Structure vs. the Frequency Code

Toni Rietveld Judith Haan Linda Heijmans Carlos Gussenhoven

Centre for Language Studies, University of Nijmegen, The Netherlands

Abstract

In this article, two hypotheses were tested to explain attitudinal ratings like SURPRISE, SUGGESTION, REMINDER etc. of four rising nuclear contours observed in a Dutch question corpus and described as (a) H*L H%, (b) H* H%, (c) L*H H% and (d) L* H%. According to one hypothesis, the middle tones in (a) and (c) should be parcelled out, such that their absence produces contours (b) and (d), respectively, predicting communality of meaning within (a, c) that excludes (b, d). That is, (b, d) could be viewed as undershot variants of (a, c), with undershoot expressing a shade of meaning different from that of the fully realised pitch accents. This hypothesis was not confirmed by the data, though. The other hypothesis was based on the concept of the 'frequency code', which associates high/rising F_0 values with non-assertive behaviour and appeared to provide a much better model to predict listeners' ratings.

Copyright © 2002 S. Karger AG, Basel

1. Introduction

1.1 Contour-final H%

Cross-linguistically, it is quite common for utterance-final pitch to revert to the speaker's baseline. That is, in the greater part of spoken utterances speakers choose a low final boundary tone (L%) whose actual F_0 value, within a given speaker, appears to be fairly constant [American English: Liberman and Pierrehumbert, 1984; German: Oppenrieder, 1988; Spanish: Prieto et al., 1996; Dutch: Kraaijeveld, 1997, 't Hart, 1998]. However, many languages additionally possess contours ending in a high boundary tone (H%), which are typically used when asking questions.

Interrogativity may be intonationally encoded in several ways. Locally, a question may exhibit a high beginning and/or a raised nuclear accent peak and/or a terminal rise. Globally, a question may be realised on a higher overall register level, and downtrends such as declination, downdrift or final lowering may be suspended [Haan, 2002]. Still, in a large number of languages a steep rise in pitch on the final syllable (H%) is considered a hallmark of questions. At the same time, however, it is also possible for final rises to occur in non-questions, signalling non-finality [Caspers, 1998].

According to Cruttenden [1994, p. 168], the dichotomy between final low/falling and final high/rising pitch is a near-universal, corresponding to meanings that can be broadly paraphrased as ‘closed’ versus ‘open’, respectively. The former term then generally covers assertive and non-continuative utterances, the latter their non-assertive and continuative counterparts.

In the view of Ohala [1983, 1984, 1994], the high/low dichotomy has a biological basis. Ohala refers to Morton [1977], who conducted a survey of the literature on the vocalisations of 28 avian and 28 mammalian species and discovered remarkable cross-species similarities in the acoustic form of the sounds used in competitive encounters. At a meeting of two potential rivals, pitch height indirectly conveys an impression of the individuals’ respective sizes, such that low pitch is associated with a large body and dominance and high pitch with a small body and submission. The use of these contrasting frequencies probably serves to avert an actual fight that might lead to injury or even death; the individual feeling least confident may express its submission by means of high(er) pitch. A familiar example is the growl of the threatening dog vis-à-vis the whine of the submissive dog. Apart from having the basic meanings ‘small’, ‘lack of threat’ and ‘subordination’, high pitch in a vocaliser may also be associated with uncertainty and dependence. Conversely, low pitch not only signals that the vocaliser is large and dominant but also, by extension, that s/he is confident and independent. Assuming that this cross-species, pre-linguistic ‘frequency code’ is an integral part of the human vocal communication system, Ohala notes that it explains a number of systematic sound-meaning correlations in human language [Chen et al., 2001]. For instance, it explains why asking a question universally involves high(er) pitch: for the desired piece of information, the speaker depends on the listener’s goodwill and co-operation. The frequency code also accounts for the observation that, cross-culturally, attitudes and emotions such as politeness or sadness are, similarly, conveyed in a non-arbitrary way by either high or low pitch [van Bezooijen, 1984].

1.2 Contour-final H% in Dutch

As far as Dutch is concerned, Keijsper [1984, p. 123] has claimed that final high/level pitch and final rising pitch both signal that ‘more is to come’. Likewise, ‘t Hart [1998, p. 103] maintains that a final rise gives an implication of ‘unfinishedness’. Such claims are supported by results of a perception experiment by Swerts et al. [1994], which made it clear that a final rise is a powerful cue to the perceived non-finality of an utterance. Obviously, these meanings fit in with Cruttenden’s [1994] notion of ‘openness’.

A production corpus of Dutch statements and questions compiled by Haan [2002] showed that utterance-final H% occurred in a large majority of question contours whereas it was absent from corresponding statements. Since the experiment reported in the present article was based on contours occurring in this particular corpus, some further information seems in order. Haan’s [2002] corpus consisted of two statements, *Renée heeft nog vlees over* ‘Renée has some meat left’ and *Marina wil haar mandoline verkopen* ‘Marina wants to sell her mandoline’, together with the corresponding wh-questions (featuring a wh-word and inversion of subject and finite verb), yes-no questions (featuring inversion) and declarative questions (lexically and syntactically identical to the statement). These four sentence types were presented in various conditions, e.g. in isolation and as members of sentence pairs, and read out twice by ten subjects (5 men, 5 women). This resulted in a corpus of 800 utterances, i.e. 200 statements,

Table 1. Relative frequencies of high (H%) and low (L%) final boundary tones in statements and corresponding question versions

Final boundary tone	Low (L%)	High (H%)
Statements (n = 200)	100.0%	–
Wh-questions (n = 200)	37.0%	63.0%
Yes-no questions (n = 200)	8.5%	91.5%
Declarative questions (n = 200)	–	100.0%

200 wh-questions, 200 yes-no questions and 200 declarative questions. Table 1 shows the distribution of L% and H% over the sentence types in percentages. Within the subset of 600 questions, there were 509 cases of final H%, which occurred in the four contours given in (1) [ToDI notation, Gussenhoven et al., 1999].

- (1) a. H*L H% ‘the fall-rise’
 b. H* H% ‘the high rise’
 c. L*H H% ‘the low rise’
 d. L* H% ‘the low low rise’

H% was most frequently preceded by H*L (306 cases). This observation is consistent with early reports by Van Es [1932] that Dutch questions, typically, display a steep accentual fall followed by a steep final rise on the last syllable. However, in an additional 89 cases such a fall was absent. Figure 1 compares two repetitions of the same utterance by speaker E.S. In figure 1a, a nuclear falling pitch accent H*L occurs, followed by H%. By contrast, in the (b) contour the L tone of H*L would seem to have been deleted. The starred accentual tone is directly adjacent to the high boundary tone, resulting in a high plateau functioning as the starting point for the rise of H%. The adjacency of H* and H%, occurring in almost a quarter out of the total number of H* accents, prompted the question of whether a similar phenomenon could be found in the relatively few utterances featuring an L*H pitch accent rather than H*, i.e. whether it was possible for L*H H% to lack its H tone and appear as L* H%. Inspection of the corpus showed that, in 40 out of the 65 occurrences, L* was directly followed by H%. Figure 2 gives the two versions. Since speakers often produced different contours for the same sentences without appearing to express different meanings, we decided to determine whether these formal differences reflected pragmatic differences by running an experiment in which listeners rated the four H%-final contours on six pragmatic scales.

2. A hypothesis: Phonological Undershoot

While it is conceivable that each of the four nuclear contours given in (1) is morphemically simplex, it seems reasonable to suppose that they are complex. One way in which they could be analysed is by identifying the H% tone as a question morpheme, leaving the four pitch accents as separate morphemes. Another, or an additional option would be to separate off L* and H* as morphemes, which analysis predicts some communality of meaning within (1a, b) that excludes (1c, d) and vice versa. Yet a further hypothesis might be that the middle tones in (1a) and (1c) should be parcelled out, such that their absence produces contours (1b) and (1d), predicting communality of meaning within (1a, c) that excludes (1b, d) and vice versa. Put differently, (1b, d) could be

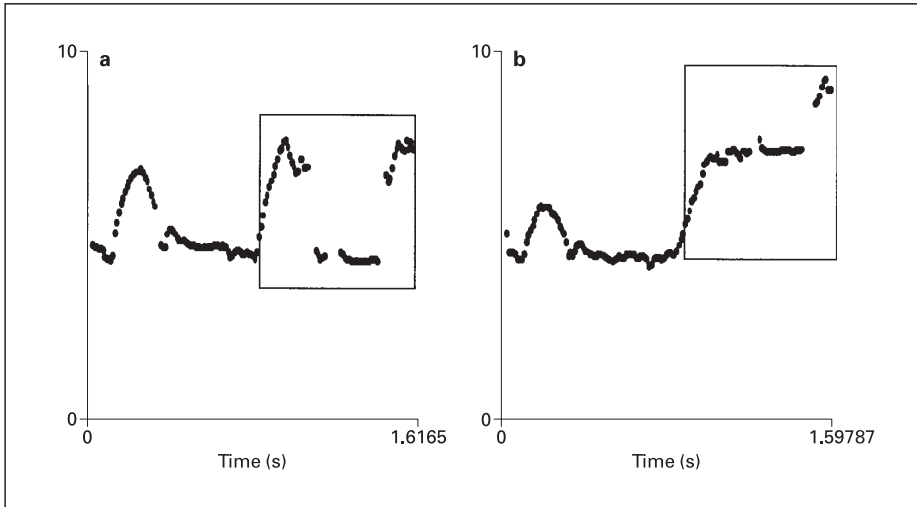


Fig. 1. Two combinations of the nuclear pitch accent H*L and the high boundary tone H% (see boxed areas) in *Marina wil haar mandoLine verkopen?* (Marina wants her mandolin (to) sell?) by speaker E.S. (female). **a** H*L H%. **b** H*H%.

viewed as undershot variants of (1a, c), with undershoot expressing a shade of meaning different from that of the fully realised pitch accents. Below, this latter hypothesis will be worked out further.

Undershoot, the failure of articulations to reach their target leading to a shortcut between the flanking targets, occurs in segmental realisations and in tonal realisation. An example of the former is Wells' [1981] 'smoothing' of British English closing diphthongs before vowels, which tends to leave the second element of a diphthong like [ai] unrealised in [fa(i)ə] *fire*. In the realisation of intonation, the L of the H*L H% contour is commonly pronounced at mid pitch when the contour is used on monosyllables in Dutch.

As a way of simplifying tonal structures, undershoot is like truncation, which affects the parts of contours at the edges of expressions. Ladd [1996, p. 133] gives some examples of languages allowing tones to remain unrealised; this commonly involves final monosyllables which have to accommodate both a pitch accent and a boundary tone. For instance, the Hungarian tonal sequence L*H L% may be produced as L*H, which is then analysed as the boundary tone L% being omitted. While these and other examples of truncation concern boundary tones, Ladd [1996] also gives an example of truncation of an *accentual* tone which, again, springs from a lack of segmental material. Thus, while in German answering the telephone on the sequence H*H% (as opposed to L*H) is considered impolite as long as a surname has at least two syllables, this is not so with a monosyllabic name like *Schmidt*. This may be explained by assuming that the latter case represents a kind of left-truncated low-rise, i.e. a sequence of L*H from which the accentual L* has been truncated [Ladd, 1996, p. 135]. Although Ladd [1996] primarily discusses truncation as a matter of phonetic realisation, he points out that it is also possible for truncation to be phonologised, and thus amount to the Undershoot of a tone from the representation.

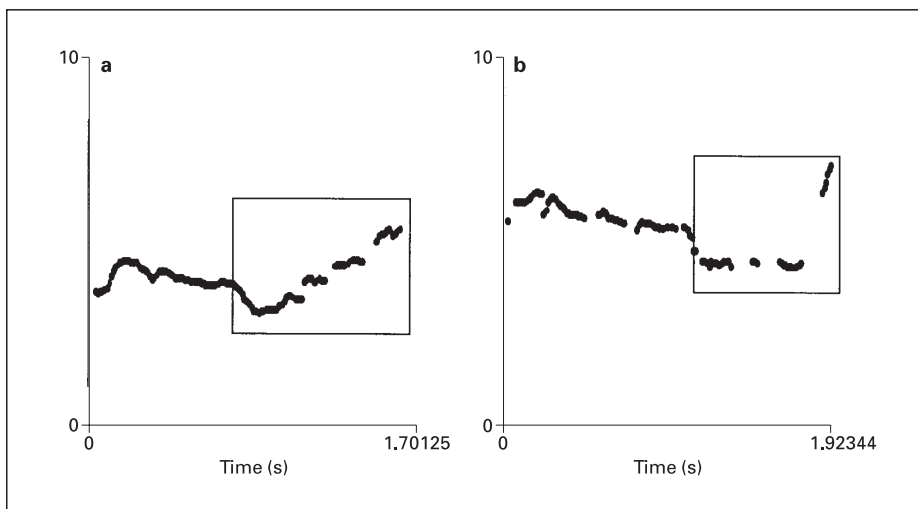


Fig. 2. Two combinations of the nuclear pitch accent L*H and the high boundary tone H% (see boxed areas): **a** L*H H% in *Marina wil haar mandoLIne verkopen?* (Marina wants her mandolin (to) sell?). Speaker H.G. (male). **b** L* H% in *Waar wil Marina haar mandoLIne verkopen?* (Where wants Marina her mandolin (to) sell?). Speaker MK (female).

However, the Dutch pattern as shown in figures 1b and 2b would not seem to be prompted by lack of segmental material. In fact, in the questions concerned as many as three unstressed syllables occur between the accented syllable and the final, unstressed syllable carrying the terminal rise. What this may mean is that an undershoot that became phonologised on monosyllables (e.g. H*L H% → H* H%) has subsequently been generalised to sentences with non-final focus. The purpose of this study was to decide whether the full and the undershot accents reflect different pragmatic choices on the part of the speaker. To this end, a perception experiment was carried out for which specific outcomes were predicted; these predictions are specified below.

2.1 Predictions

If undershoot is an intonational morpheme expressing a distinct meaning, one would expect it to affect falling and rising accents in the same manner. That is, if judgements turn out to systematically differ as a function of full contour versus undershot contour, they should do so for both pitch accents and in the same direction, as illustrated by figure 3. On the other hand, if full and undershot contours do not appear to differ in comparable manners, the hypothesis that there is a common morpheme ‘undershoot’ will not be supported. Some other pairing might be suggested by the data, leading to other hypotheses (fig. 4), or alternatively, no pairing may emerge, so that each of the four contours would be best seen as constituting a separate, simplex, morpheme by itself.

Nuclear accent + high boundary tone	FULL CONTOUR		UNDERSHOT CONTOUR
H*L H%	H*L H%	↔	H* H%
L*H H%	L*H H%		L* H%

3

Nuclear accent + high boundary tone	FULL CONTOUR		UNDERSHOT CONTOUR
H*L H%	H*L H%	↕	H* H%
L*H H%	L*H H%		L* H%

4

Fig. 3. The effect of tonal undershoot if undershoot is an intonational morpheme.

Fig. 4. The effect of tonal undershoot if undershoot is not an intonational morpheme.

3.1 The experiment

3.1.1 Materials

Varying the two pitch accents preceding the high final boundary tone H% [the PSOLA procedure implemented in PRAAT was used, Boersma and Weenink, 1996] yielded the four contours shown in figures 1 and 2. Moreover, as an earlier experiment showed that choice of onset (pre-head) affects judgements of the subsequent pitch accent [Grabe et al., 1997], the four contours were crossed with two pre-heads, high vs. low. The eight experimental contours are presented in figure 5. Two female speakers of Dutch realised the contours on four one-accent prepositional phrases referring to Dutch towns or villages (with the accent always on the first syllable of a trisyllabic place-name): (a) *op de weg naar Almelo* ('on the road to Almelo'); (b) *heel dicht bij Harlingen* ('quite close to Harlingen'); (c) *halverwege Millingen* ('halfway to Millingen'), and (d) *iets voorbij Hengelo* ('somewhat beyond Hengelo'). These phrases were chosen because they were assumed to lack connotations which might bias listeners' judgements. Six pragmatic scales were selected which were felt to combine well with the 'openness' signalled by a high final boundary tone. They were presented on the score forms in the following way (in parentheses we give the labels we are going to use throughout this article when referring to these scales):

'By her intonation, the speaker indicates that...

- (1) ... she expects an answer (yes or no) from the listener' (QUESTION)
- (2) ... she is surprised about the contents' (SURPRISE)
- (3) ... the utterance is meant as a suggestion' (SUGGESTION)
- (4) ... the utterance is meant as a challenge to the listener' (CHALLENGE)
- (5) ... she wishes to remind the listener of something' (REMINDER)
- (6) ... she asks the listener to agree' (APPEAL)

3.2 Recording and Test Procedures

Recordings of the test utterances were made in a single session in a soundproof booth. Previously, the 2 speakers had had some practice as some contours came more naturally to them than others. They were provided with the 32 written utterances, together with a visual representation of the con-

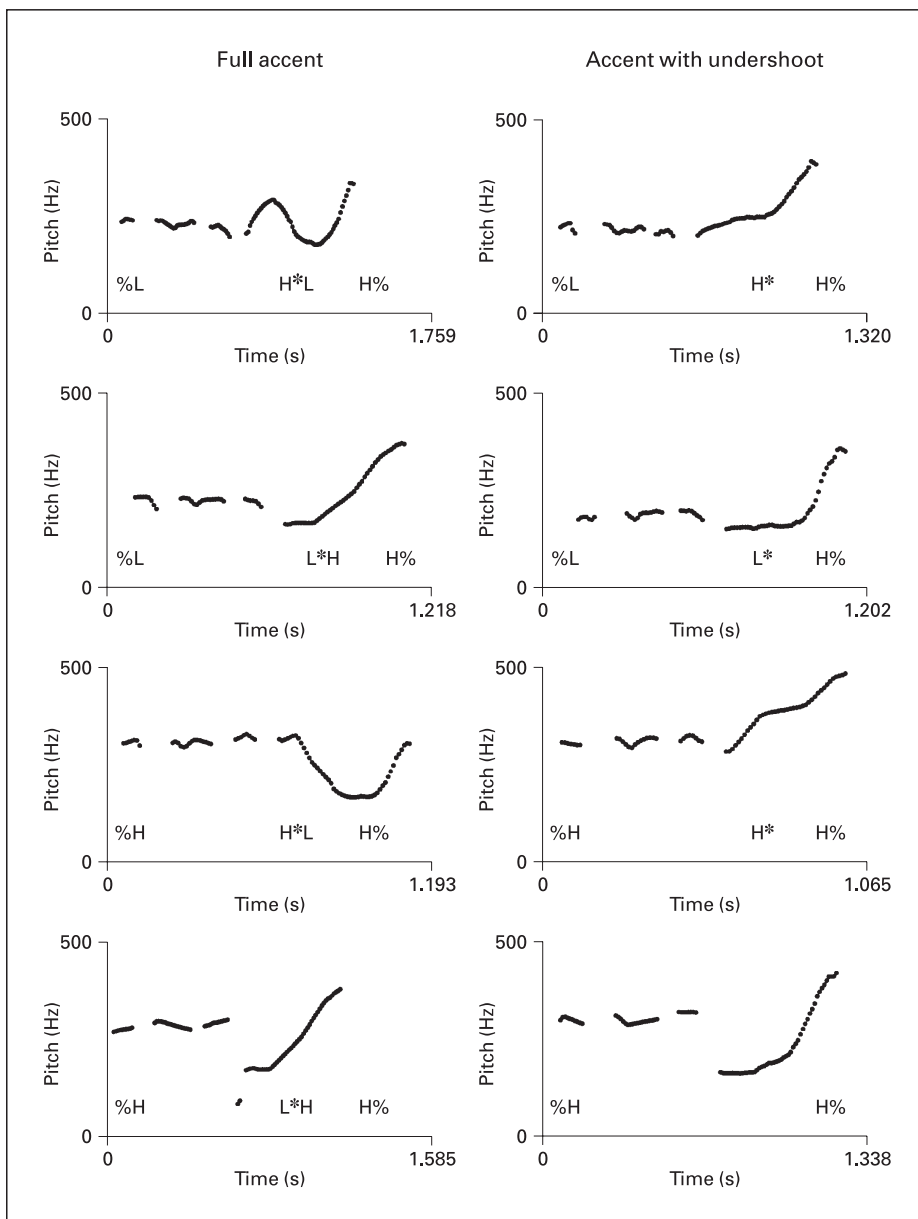


Fig. 5. The eight experimental pitch contours.

Table 2. Results of the analyses of variance on the judgements on the six scales

Factor	Scale					
	QUESTION	SURPRISE	SUGGESTION	CHALLENGE	REMINDER	APPEAL
Sentence (S)	*	*	*	*	*	–
Onset (O)	*	*	*	–	*	*
Accent (A)	*	–	*	–	–	*
Undershoot (U)	*	*	–	–	*	*
S × O	*	*	*	–	–	*
S × A	–	*	–	–	*	*
S × U	*	–	–	–	–	–
A × O	*	–	–	–	–	*
O × U	*	*	–	–	*	*
A × U	*	*	*	–	*	*
S × O × A	*	*	*	–	*	*
S × O × U	*	*	–	–	*	–
S × A × U	*	*	–	–	–	*
O × A × U	–	–	*	–	*	–
S × O × A × U	–	–	*	–	*	–

Significant effects (1% level) are marked by *, non-significant results by –.

tours. Their performance was being monitored by the second author and, if necessary, certain tokens were repeated till the required contour was produced. The utterances of the speaker who performed best were analysed with the PRAAT package [Boersma and Weenink, 1996]. The extracted F_0 contours were checked to see whether the F_0 of H^* and $H\%$ fell within a predetermined 10-Hz range; if they did not, they were adjusted accordingly. This ensured that all utterances had approximately the same pitch range. Six tapes were made, each of which presented the 32 stimuli in a different random order (sampling frequency 16 kHz). Thirty-three native speakers of Dutch, recruited from the student population, took part in the listening experiment: 13 male and 20 female students, with ages ranging between 19 and 36 years. They were divided into three groups for each of which the order of the six tapes was systematically varied, as was the order of the six pragmatic scales. Varying the order of presentation of the stimuli at several levels was assumed to preclude order effects on the scores.

Subjects were asked to imagine that the utterances were spoken by a drama student whose performance they had to assess. That is, given the pragmatic scale on the score form, how appropriate was the contour the speaker produced? Each stimulus was played once and the subjects had to enter their judgements on a ten-point scale, as is customary in Dutch education tests. Thus, ‘1’ stood for ‘the intonation contour does not match the given pragmatic scale at all’, whereas ‘10’ meant ‘the intonation contour perfectly matches the given pragmatic scale’. The three groups of subjects listened to the stimuli over headphones; it took them 30 minutes to complete the test, for which they were paid a fee.

4. Results

4.1 Analysis of the Judgements

The judgements on the six separate scales were submitted to analyses of variance (SPSS: MANOVA), with repeated measures on four fixed within-subject factors, Sentence (four utterances), Onset (high and low), Accent (H^* or L^*) and Undershoot (with or without). In table 2 we give an overview of the results of the analyses; Huynh-Feldt adjusted p values [Rietveld and van Hout, 1993] were used where necessary.

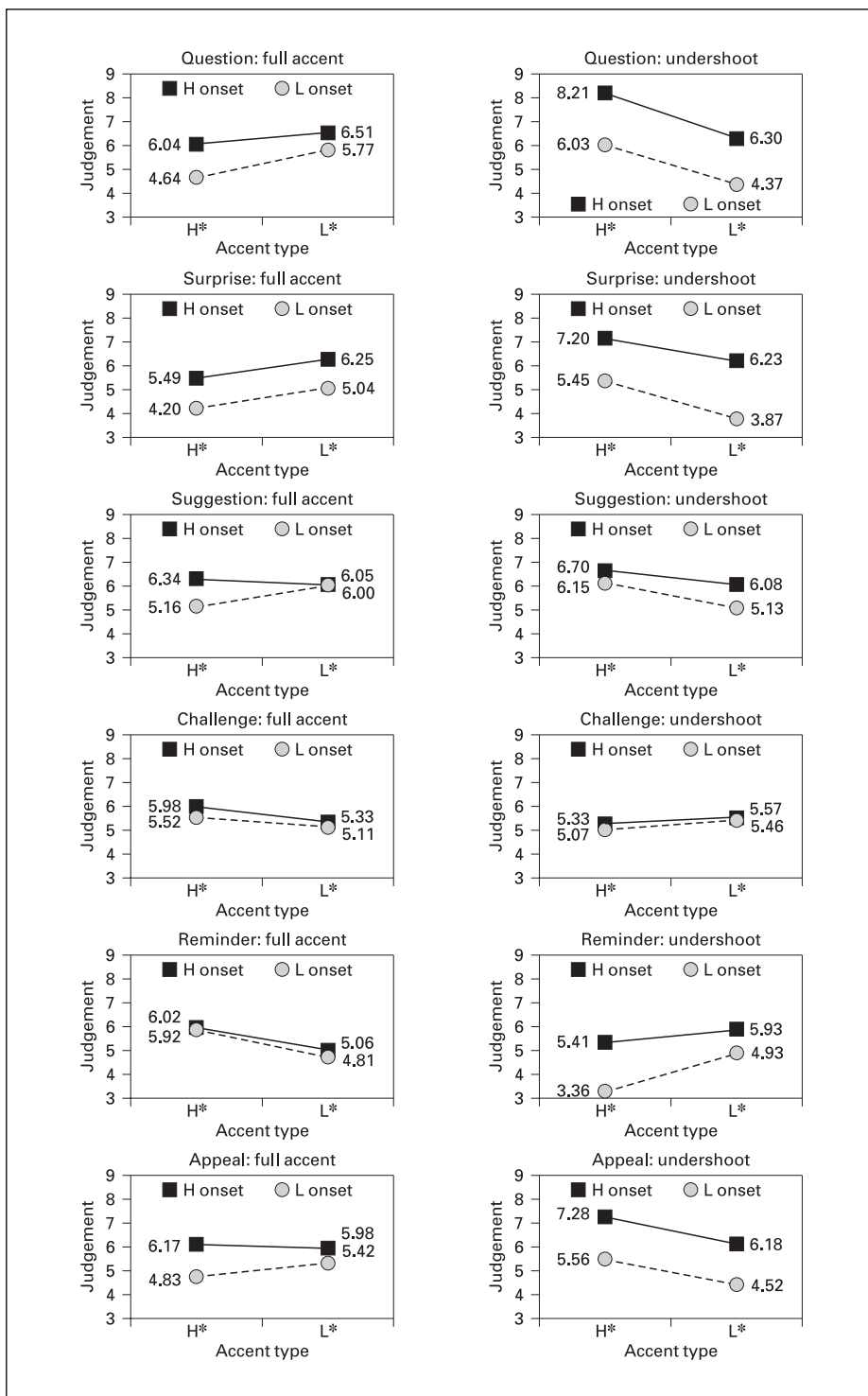


Figure 6 shows the effect of undershoot on the scores as a function of pitch accent type and onset. On the scales SURPRISE, QUESTION, APPEAL and SUGGESTION, undershoot yields higher scores when combined with the H*L accent; in REMINDER and CHALLENGE, it is the other way round. Undershoot is a relatively strong factor (the measure of effect size used here is partial η^2) for the scales QUESTION ($\eta^2 = 0.40$), SURPRISE ($\eta^2 = 0.38$) and CHALLENGE ($\eta^2 = 0.45$), but has a much smaller effect for the scales APPEAL ($\eta^2 = 0.12$), SUGGESTION ($\eta^2 = 0.04$), and REMINDER ($\eta^2 = 0.03$).

The results can be summarised as follows: (1) Interactions with Sentence are all ordinal; this means that the results on the four sentences all show similar tendencies. (2) %H onsets give rise to more positive judgements than %L onsets, but for one scale: CHALLENGE. (3) Undershoot with H*accents gives rise to more positive judgements while with L*accents it yields more negative judgements (exceptions: CHALLENGE and REMINDER).

In view of the significant interactions Undershoot by Accent for nearly all scales (table 2), simple main effects analyses were carried out in order to find out to what extent the factor undershoot had similar effects for the two accents at issue. As explained above, if it does, undershoot could be seen as a common ‘morpheme’ with a distinct meaning, but if it does not, undershot accents reflect different pragmatic choices of the speaker, depending on the accent. These analyses were performed for high and low onsets separately at a significance level of 1% ($df_1 = 1$, $df_2 = 32$) (table 3).

The quite large number of simple main effects can be summarized as follows: the factor ‘undershoot’ has always an effect on scale values for QUESTION and SURPRISE when associated with an H* accent. For L* accents the situation is somewhat more complex. With a high onset, ‘undershoot’ is never significant; with a low onset, however, it always is except for the scale CHALLENGE.

4.2 Discussion 1

Clearly, the results did not confirm the hypothesis that ‘undershoot’ is a morpheme with a distinct meaning, since its effect depends on the specific accent to which it applies (fig. 3), and although there would seem to be slightly stronger evidence for the alternative hypothesis, i.e. that undershoot has different effects with H* accents than with L* accents (fig. 4), the results here were not clear-cut. As a matter of fact, when viewed from another angle the results suggested an entirely different approach, one which takes pitch levels into account, rather than morphological structure. Obviously, undershoot affected the proportion of high pitch in the experimental material in that it *increased* the amount of high pitch in the nuclear contours with H*L accents and *decreased* the amount of high pitch in contours with L*H accents. According to the frequency code proposed by Ohala [1983, 1984, 1994] (cf. section 1.1), relatively low pitch universally reflects a speaker’s confidence and authority, whereas relatively high pitch signals that the speaker is somehow dependent on the listener. It was found that on the scales which are, intuitively, associated most with ‘dependence’ (i.e. QUESTION, SURPRISE and APPEAL) the highest scores were elicited by the contours with a maximal proportion of high pitch, i.e. %H H* H% (with an undershot accent). Here mean scores

Fig. 6. The effect of undershoot on the scores as a function of pitch accent and onset.

Table 3. Results of simple main effects analyses of factor Undershoot at the two levels of the factor Accent (H* and L*), for the different onsets (%H and %L)

Scale	H*	L*
<i>High onset (%H)</i>		
QUESTION	*	–
SURPRISE	*	–
SUGGESTION	–	–
CHALLENGE	–	–
REMINDER	*	–
APPEAL	*	–
<i>Low onset (%L)</i>		
QUESTION	*	*
SURPRISE	*	*
SUGGESTION	*	*
CHALLENGE	–	–
REMINDER	–	*
APPEAL	*	*

Significant level set at 1%. Significant effects are marked by *, non-significant results by –.

were 8.21 (QUESTION), 7.20 (SURPRISE) and 7.28 (APPEAL). By contrast, pitch was maximally low in the %L L* H% contours (with an undershot accent). On the same scales, these low contours yielded the considerably lower scores of 4.37 (QUESTION), 3.87 (SURPRISE) and 4.52 (APPEAL).

Conversely, on the scales which are associated with ‘self-confidence’ and ‘authority’, i.e. CHALLENGE and REMINDER, the scores for the maximally high contours were only 5.33 and 5.41, respectively. As for SUGGESTION, scores on this scale seemed to occupy an intermediate position which, likewise, can be explained in terms of the frequency code. The meaning ‘suggestion’ impinges on the frequency scale in two, contradictory ways. On the one hand, it is an assertion rather than a request for information, and as such is associated with the lower end of the frequency scale. On the other hand, as an assertion it is only mildly enforced, and as such contains a meaning component associated with the higher, ‘friendlier’ end of the frequency code. The pitch variation will be perceived in terms of varying degrees of friendliness. In other words, the results seemed to conform to a pattern predicted by Ohala’s frequency code, to the extent that high pitch is interpreted by listeners as signalling a speaker’s dependence, whereas low pitch indicates finality, authority and assurance. Table 4 gives the mean scores on the maximally high and maximally low contours (the results of undershoot). The alternative hypothesis – introduced in section 1.2 – that the experimental results actually had their basis in Ohala’s frequency code rather than in differences in morphological structure was then put to the test; results are reported in the next section.

4.3 Testing the Frequency Code: Method and Results

To test the explanatory power of the frequency code, we created one new within-subject variable, Pitch, with two levels, the highest contour (%HH* H%) and the lowest (%LL* LH%). The judgements on the six separate scales were submitted to analyses of variance (SPSS: MANOVA), with repeated measures on two fixed within-subject factors, Pitch (two levels) and Sentence (four utterances). The analyses were

Table 4. Mean scores on the six scales as a function of the two contours with maximally high (left column) and maximally low (right column) pitch

Scale	%HH* H%	%LL* H%
QUESTION	8.21	4.37
SURPRISE	7.20	3.87
APPEAL	7.28	4.52
SUGGESTION	6.70	5.13
CHALLENGE	5.33	5.46
REMINDER	5.41	3.36

used to test the assumption that a sequence of tonal segments which yields a maximally high pitch contour (%HH* H%) yields significantly higher ‘appropriate-ness’ judgements for the scales SURPRISE, QUESTION, APPEAL and SUGGESTION but not for the scales CHALLENGE and REMINDER. This assumption was confirmed in the analyses. The factor Pitch was significant for the four scales mentioned above: SURPRISE: $F(1, 32) = 96.90$, $p < 0.01$, $\eta^2 = 0.75$, QUESTION: $F(1, 32) = 111.20$, $p < 0.01$, $\eta^2 = 0.78$, APPEAL: $F(1, 32) = 39.89$, $p < 0.01$, $\eta^2 = 0.55$, and SUGGESTION: $F(1, 32) = 11.31$, $p < 0.01$, $\eta^2 = 0.26$. No significance was reached for the scale CHALLENGE ($F(1, 32) = 0.90$, $p = 0.351$, $\eta^2 = 0.03$), and for the scale REMINDER the lowest contour (%LL* H%) was judged most appropriate: $F(1, 32) = 42.92$, $p < 0.01$, $\eta^2 = 0.57$. There were no significant interactions Pitch by Sentence at the 0.01-level but for the scale SURPRISE. This interaction, however, was ordinal. It is interesting to see that the effect size (expressed by the index η^2) is highest for QUESTION, which gave rise to the hypotheses tested in the present investigation (0.78), and decreases via SURPRISE (0.75), APPEAL (0.56), SUGGESTION (0.26) to the lowest, obtained for CHALLENGE (0.03), the scale which is possibly most remote from the meaning of ‘question’.

In order to cross-validate the results of the analyses of variance, we also approached the hypothesis of the frequency code more directly. To that end we calculated the average pitch values (expressed in Hertz) for five parts of each utterance: ‘onset’ (the syllables before the accented syllable), ‘antepenultimate (S1)’, ‘penultimate (S2)’ and ‘last syllable (S3)’ of each utterance; in addition, the temporally weighted averages of these four pitch values were calculated: ‘Total’. In table 5 we give the values of the correlation coefficients between the average rating values (pooled across all subjects) for each of the scales and the five parts.

4.4 Discussion 2

The first question to be answered by this study was whether, in utterances ending in H%, the formal difference between full accent and accent with undershoot should be interpreted as a choice between unrelated morphological categories (i.e. H*L, H*, L*H, and L*) or whether there is a morpheme ‘undershoot’ which introduces a common meaningful element to the H*L and the L*H pitch accents. In the latter case, judgements ought to differ as a function of full vs. undershot accent for both H*L and L*H, and in the same direction. In the previous sections, we established that the middle tone does not express a common meaning component. What our experiment did reveal was that pitch level was interpreted gradiently: higher or lower pitch levels appeared to

Table 5. Pearson correlations between mean rating values and average pitch (in Hertz) on onset, syllables S1, S2 and S3 and the temporally weighted average of these values: total

Scale	Part of utterance				
	onset	S1	S2	S3	total
QUESTION	0.71*	0.53*	0.68*	0.79*	0.91*
SURPRISE	0.75*	0.41	0.56*	0.71*	0.87*
SUGGESTION	0.65*	0.52*	0.51*	0.59*	0.79*
APPEAL	0.80*	0.60*	0.61*	0.64*	0.92*
REMINDER	-0.49*	-0.32	-0.64*	-0.87*	-0.76*
CHALLENGE	0.10	0.10	-0.36	-0.68*	-0.21

Critical value of correlations at the 1% level (two-sided) = $|0.45|$; $n = 32$.

be interpreted in terms of Ohala's frequency code, independently of the presumed intrinsic meanings of the H*L and L*H pitch accent. On the basis of this code, we expected that scales reflecting a speaker's dependence, submissiveness, etc. would yield high scores for stimuli with high pitch levels (i.e. positive correlations) and conversely, that scales reflecting a speaker's authority would produce high scores for stimuli with low pitch levels (i.e. negative correlations). This expectation was generally borne out by the results, although the results on the scale CHALLENGE were less clear-cut (see table 5).

The average absolute value of the correlation coefficients between the total average pitch and the mean ratings on all scales but for CHALLENGE was high: $|0.85|$. This result strongly supports the hypothesis that listeners used the frequency code when interpreting the meanings of all contours. This conclusion is strengthened by the finding that a negative correlation (-0.76) was found between average pitch and scores on the scale REMINDER, since low pitches are associated with meanings like certainty and assertiveness, both of which are consonant with the act of reminding someone. The clarity of these results leaves little room for alternative interpretations.

5. Summary and Conclusions

In a production experiment reported by Haan [2002] involving a reading task carried out by 10 native speakers of Dutch, four H%-final contours were used on questions, H*L H%, H* H%, L*H H% and L* H%. Although there seemed to be a high degree of variation in the distribution of the contours, it seemed difficult to believe that four clearly distinguishable contours were simply variant forms expressing the same morpheme. On the other hand, if there is a high degree of interchangeability, it might be that there are subsets of semantically related contours. In particular, we hypothesised that deletion of the middle tone (the 'trailing' tone of H*L and L*H) might have an invariant pragmatic effect, and that H*LH% was semantically related to H*H% in a way that was comparable to the relation between L*HH% and L* H%. In an attempt to further support this hypothesis, we included both high-onset and low-onset versions of the four contours, hoping to find a near-invariant contribution of the 'trailing' tone across the four contour pairs. The results of the experiment did not support this hypothesis. On the contrary, on

most scales the effect of the deletion of the trailing L tone was opposite to that of the deletion of the trailing H-tone. That is, judgements did not vary as a function of full vs. undershot versions of the L*HH% and H*LH% contours. The presence and absence of undershoot in the two contours appear to have affected ‘universal meaning’ coded in the pitch contours. In the case at hand, undershoot of L from the H*L accent preceding H% causes pitch to retain its high level. Conversely, undershoot of H from the L*H accent preceding H% results in a prolongation of the low pitch level. As was seen above, in the frequency code, high pitch levels universally signal deference, uncertainty etc. while, conversely, low pitch levels reflect confidence, authority and related attitudes. Therefore, one would predict that undershoot of the H*L accent results in higher scores on scales representing uncertainty, whereas undershoot of the L*H accent can be expected to produce higher scores on scales reflecting a speaker’s confidence and assertiveness.

The high positive correlations between average pitch, measured in different parts of the contour, and scores on the four ‘uncertainty’ scales QUESTION, SURPRISE, APPEAL and SUGGESTION, together with the negative correlation between average pitch and scores on REMINDER suggest that the frequency code is by far the more important element in an explanation of our subjects’ scores. That is, subjects would appear to have judged the presence vs. absence of undershoot in the nuclear contours mainly in terms of Ohala’s frequency code.

A number of conclusions can be drawn. Most evidently, our experiment lends unsought support for Ohala’s [1983, 1984, 1994] proposal that the interpretation of high vs. low pitch can be guided by a code according to which high pitch signals meanings associated with a small vocaliser (uncertainty, vulnerability, dependence, etc.) and low pitch with a large vocaliser (certainty, protectiveness, dominance, etc.) Second, we found no support for any clustering of phonological contours of the sort that was suggested by our hypothesis, and an analysis whereby the trailing tone in a T* – T% context represents a morpheme (the Undershoot analysis) could not be supported. Third, pragmatic meanings that are closely associated with the meanings embodied in the frequency code will indeed trigger that code. It is possible that the experimental setting, in which judges had to judge series of isolated utterances for pragmatic meanings is particularly likely to elicit interpretations of this kind, interpretations which are probably due to a language-independent, pre-linguistic interpretation strategy. Arguably, this effect is so strong that any contribution of the structural meanings of intonational morphemes may well be masked. This result suggests that for the investigation of the functions of structural properties of intonation contours, more sophisticated methodologies than the one used in this investigation may be called for.

References

- Bezooijen, van R.: Characteristics and recognizability of vocal expressions of emotion (Foris, Dordrecht-Riverton 1984).
- Boersma, P.; Weenink, D.: PRAAT: A system for doing phonetics by computer. Rep. Inst. Phonet. Sci., University of Amsterdam, Vol. 132 (1996).
- Caspers, J.: Who’s next? The melodic marking of question versus continuation in Dutch. *Lang. Speech* 41: 375–398 (1998).
- Chen, A.; Rietveld, T.; Gussenhoven, C.: Language-specific effects of pitch range on the perception of universal intonational meaning; in Proc. Eurospeech 2001, Aalborg, 2001, vol. 2, pp. 1403–1407.
- Cruttenden, A.: Rises in English; in Windsor Lewis, *Studies in general and English phonetics: essays in honour of Professor J.D. O’Connor*, pp. 155–173 (Routledge, London 1994).

- Grabe, E.; Gussenhoven, C.; Haan, J.; Marsi E.; Post B.: Preaccental pitch and speaker attitude in Dutch. *Lang. Speech 41*: 63–85 (1997).
- Gussenhoven, C.; Rietveld, T.; Terken, J.: ToDI. <http://lands.let.kun.nl/todi> (1999).
- Haan, J.: Speaking of questions. an exploration of Dutch question intonation; PhD thesis University of Nijmegen (2002).
- 't Hart, J.: Intonation in Dutch; in Hirst, Di Cristo, *Intonation Systems*, pp. 96–111 (Cambridge University Press, Cambridge 1998).
- Keijsper, C.E.: Vorm en betekenis in Nederlandse toonhoogtecontouren, parts 1 and 2. *Forum der Letteren 25*: 20–37, 113–126 (1984).
- Kraaijeveld, J.: Idiosyncrasy in prosody: speaker and speaker group identification; PhD thesis University of Nijmegen (unpublished, 1997).
- Ladd, D.R.: *Intonational phonology* (Cambridge University Press, Cambridge 1996).
- Lieberman, M.; Pierrehumbert J.: Intonational invariance under changes in pitch range and length; in Aronoff, Oehrle, *Language and sound structure*, pp. 157–233 (MIT, Cambridge 1984).
- Morton, E.W.: On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *Amer. Naturalist 111*: 855–869 (1977).
- Ohala, J.J.: Cross-language use of pitch: an ethological view. *Phonetica 40*: 1–18 (1983).
- Ohala, J.J.: An ethological perspective on common cross-language utilization of F₀ of voice. *Phonetica 41*: 1–16 (1984).
- Ohala, J.J.: The frequency code underlines the sound symbolic use of voice of pitch; in Hinton, Nichols, Ohala, *Sound symbolism*, pp. 325–347 (Cambridge University Press, Cambridge 1994).
- Oppenrieder, W.: Intonatorische Kennzeichnung von Satzmodi; in Altmann, *Intonationsforschungen* (Niemeyer, Tübingen 1988)
- Prieto, P.; Shih, C.; Nibert H.: Pitch downtrend in Spanish. *J. Phonet. 24*: 445–473 (1996).
- Rietveld, T.; van Hout, R.: Statistical techniques for the study of language and language behaviour (Mouton de Gruyter, Berlin 1993).
- Swerts, M.; Bouwhuis, D.G.; Collier, R.: Melodic cues to 'finality' in discourse units. *J. acoust. Soc. Am. 96*: 2064–2075 (1994).
- Van Es, G.A.: *Syntaxis en dialectstudie: intonatie en syntaxis*, parts 1, 2, 3. *Onze Taaltuin 1*: 87–94, 122–128, 169–175 (1932).
- Wells, J.C.: *Accents of English* (Longman, London 1981).